Conjunctive Markers of Contrast in English and French

From syntax to lexis and discourse

Maïté Dupont

JOHN BENJAMINS PUBLISHING COMPANY

Conjunctive Markers of Contrast in English and French

Studies in Corpus Linguistics (SCL)

ISSN 1388-0373

SCL focuses on the use of corpora throughout language study, the development of a quantitative approach to linguistics, the design and use of new tools for processing language texts, and the theoretical implications of a data-rich discipline.

For an overview of all books published in this series, please see
benjamins.com/catalog/scl

Volume 99

Conjunctive Markers of Contrast in English and French
From syntax to lexis and discourse
by Maïté Dupont

Conjunctive Markers of Contrast in English and French

From syntax to lexis and discourse

Maïté Dupont

Université catholique de Louvain

John Benjamins Publishing Company

Amsterdam / Philadelphia

Table of contents

List of tables

List of figures

List of abbreviations

CA(s)	Conjunctive adjunct(s)
CART	Classification and Regression Tree
CCR	Cognitive approach to Coherence Relations
CFA	Contrastive Functional Analysis
CL	Corpus linguistics
CM(s)	Conjunctive marker(s)
df	Degree of freedom
FSP	Functional Sentence Perspective
LOCRA	Louvain Corpus of Research Articles
Mult-Ed	Multilingual Corpus of Editorials
p	p-value
PDTB	Penn Discourse Treebank
pmw	Per million words
RST	Rhetorical Structure Theory
SDRT	Segmented Discourse Representation Theory
SFL	Systemic Functional Linguistics
SVO language	'Subject-Verb-Object' language
TC	*Tertium comparationis*
V2 language	Verb-second language
WST	*WordSmith Tools*

Acknowledgements

This book is the result of ten years of interest in corpus-based contrastive linguistics and discourse analysis. I would like to thank wholeheartedly all the people without whom this project would never have come to fruition, whether they be colleagues, friends or family – although at this stage, the boundary between these three categories feels nearly as fuzzy as that between contrast and temporality.

First of all, I am greatly indebted to my mentors, Profs. Sylviane Granger and Liesbeth Degand. Ever since I was a bachelor student, Sylviane has been pushing me to tap into potential that I never knew I had. I am immensely grateful to her for her unshakeable confidence, her unwavering support and her endless enthusiasm towards my work. The present book owes a great deal to her invaluable input and investment. Despite her busy schedule, Liesbeth has always made herself available for – most commonly unplanned – talks about my work in her office. I was also very lucky to benefit from her extensive expertise and her vast network of relationships in the field of discourse analysis, which were a great asset in the framework of the research presented in this book.

I also wish to express my warmest thanks to Prof. Gaëtanelle Gilquin (and her mind-blowing analytical capabilities), Prof. Marie-Aude Lefer (also known as my 'contrastive linguistics godmother'), Prof. Hilde Hasselgård and Prof. Sandrine Zufferey (who were both kind enough to welcome me on extremely enriching research stays while I was working on this book) for all the time and consideration that they devoted to my work. I would also like to thank Dr. Magali Paquot, Alain Guillet and Jonathan Dedonder, who were always willing to answer the (numerous) questions that I had about statistics.

Heartfelt thanks are also owed to my dear colleagues from the Centre for English Corpus Linguistics, who created the best working environment I could ever have dreamed of, characterised by positive emulation, mutual support and benevolence. Some of my best friendships were born there, and I will forever cherish my time as a CECL member. I also acknowledge the financial support of the Belgian National Fund for Scientific Research (F.R.S.-FNRS).

Finally, I wish to thank my family and friends, who have always supported me and believed in me, while also forcing me to maintain a healthy work-life balance by taking me out for drinks, barbecues and cupcakes. Last but not least, a million thanks to my official team of supporters, Martin and Félicie, for letting me do what I like doing, however demanding it can be.

Introduction

1.1 Background

Conjunctive markers are linguistic items whose main function is to signal explicitly a logical relationship (such as addition, causality or contrast) between two clauses, sentences or paragraphs, thereby contributing to textual cohesion. They form a broad functional category grouping members from various grammatical classes, viz. adverbs (e.g. *therefore, however*), coordinators (e.g. *and, but*), subordinators (e.g. *because, although*) and various fixed phrases (*on the other hand, au contraire*). One important feature of these items is that they are (often) optional: in many contexts where an explicit conjunctive marker is used, writers and speakers can also choose to resort to the mere juxtaposition of the discourse segments, as the type of relationships expressed by conjunctive markers can usually be inferred even in their absence (Taboada 2006; Das & Taboada 2013; Zufferey 2016). In Example (1) below, for instance, it is relatively easy to understand that the two sentences are united through causality, despite the absence of an explicit causal marker such as *because*. Likewise, in (2), it is clear that a relation of contrast – which could be expressed by markers such as *by contrast* or *whereas* – holds between the two sentences in the example.

(1) She didn't come to the party. She was ill.

(2) John likes coffee. Eliza prefers tea.

While most languages of the world have a range of devices at their disposal to signal the logical relations that hold between two segments of discourse, research has shown that, even when they are typologically close, languages differ not only (i) in terms of the types of conjunctive devices that they tend to prefer to express the relations that hold between two discourse units; but also (ii) in the degree to which they need to signal such logico-semantic relationships explicitly in text (see e.g. James 1980: 113; Chesterman 1998: 185; Dixon & Aikhenvald 2011 on this). The extent of these cross-linguistic differences is partly reflected in the challenges that conjunctive markers pose for language learners and translators alike. For example, a very large number of studies have underlined the difficulties encountered by

learners of English to attain target-like conjunctive marker usage, not only regarding frequencies of use (i.e. patterns of over- and underuse), but also with respect to the semantic, syntactic or stylistic features of these markers (i.e. patterns of misuse; cf. e.g. Granger & Tyson 1996; Altenberg & Tapper 1998; Bolton et al. 2002; Tankó 2004; Rørvik & Egan 2013). One of the reasons advanced by these researchers to account for such difficulties pertains to language transfer, resulting from the differences that exist between the learners' mother and target languages. Likewise, a number of studies have underlined the problems posed by conjunctive markers for translators: for example, some scholars have uncovered significant differences in the frequency and/or types of markers used in original and translated texts from the same language (e.g. Xiao & Yue 2009; Xiao 2010a; Cartoni et al. 2011; Granger 2018; see also Halverson 2004 for an overview of the 'translation problems' posed by conjunctive markers). According to Halverson (2004: 569), some of the non-target-like uses uncovered in translated texts may be ascribed to the "obvious cross-linguistic variation in connective use". These elements make clear the substantial benefits that can be gained from the contrastive analysis of conjunctive markers: by pinpointing the ways in which the discourse organisation strategies of two languages differ, researchers may formulate valuable advice for both learners and translators. In summary, as argued by Schmied (2009: 1149), although "coherence is a feature of texts in all languages, […] to what extent formal devices of cohesion are used to signal a universal phenomenon deserves close comparative study".

With respect to the English-French language pair, the dominant claim found in the contrastive literature is that French tends to use more explicit conjunctive markers than English, which is said to have a tendency to leave the logical connections between clauses and sentences for the reader (or listener) to infer. According to Hervey and Higgins (1992: 49), for example, "it is more common in French than in English for texts to be explicitly structured with connectors ('or', 'donc', 'ainsi', 'en effet', 'par ailleurs', 'en revanche', etc.) that signpost the logical relationships between sentences" (see also e.g. Vinay & Darbelnet 1995: 234; Poncharal 2005). Accordingly, translation manuals frequently instruct their readers to add explicit conjunctive markers when translating from English into French and remove some of them when working in the other translation direction. In addition, some contrastive linguists have argued that, when they are in fact signalled explicitly in both languages, discourse relations are most typically expressed by means of coordinators in English, whereas French tends to prefer subordinators and conjunctive adjuncts – a category of link words which includes not only adverbs and adverb phrases (e.g. *however, even so*) but also prepositional phrases such as *in contrast* or *on the other hand* (e.g. Guillemin-Flescher 1981: 143;

Chuquet & Paillard 1987: 151; Vinay & Darbelnet 1995: 238).[1] The problem with these studies, however, is that they suffer from a severe lack of empirical foundation: they are either introspection-based, or rely on very small amounts of data – usually restricted to a single text type – to make very general assertions about the differences between the languages as wholes. They do not offer any quantifications of the phenomena observed but typically provide limited sets of hand-picked examples (such as the apparent lack of direct English equivalents of the French markers *en effet* and *or*) as sole evidence for their claims. The lack of generalisability of the conclusions drawn in these studies is reflected by the fact that a small number of researchers have also made contradictory statements to those described above, claiming that it is in fact English that requires discourse relations to be marked by explicit conjunctive markers, whereas French displays a tendency to simply juxtapose clauses and sentences without specifying the relations between them (see Guillemin-Flescher 1981: 83; Gallagher 1995; Mason 1998). In summary, to this day, the contrastive evidence concerning the general discourse organisational strategies of English and French remains largely inconclusive, and there is still ample scope for a rigorous study comparing the frequency and types of markers used in these two languages based on a large empirical foundation. Importantly, this is not to say that English and French conjunctive markers have *never* been the object of rigorous corpus-based contrastive studies. However, the corpus research currently available mostly consists of fairly focused studies aiming to assess the degree of cross-linguistic equivalence between the members of a few pairs of markers (see e.g. Lewis 2006 on the degree of semantic correspondence between the markers *on the contrary/au contraire* and *on the other hand/en revanche* in English and French). While such studies provide invaluable information on the semantic specificities of the markers investigated, they only offer limited insights into more general differences regarding the degree of cohesive explicitness or the preferred types of conjunctive devices used in the two languages.

1.2   Objectives of the study

The overarching objective of the present study is to compare the discourse organisation strategies of written English and French on a solid empirical basis. The study focuses on one logico-semantic relationship in particular, viz. contrast, defined as

---

1.   Conjunctive adjuncts are also called 'linking adverbials' (e.g. Biber et al. 1999) or 'conjuncts' (e.g. Quirk et al. 1985).

a broad semantic category subsuming several relations which have in common that they imply a negation of either similarity, implication or validity between two events or pieces of information. Contrast may be expressed by such markers as *but, however, whereas, on the contrary; mais, cependant, alors que, plutôt*, etc. In this study, English and French conjunctive markers of contrast are approached from an onomasiological perspective, which means that I attempt to include all the items that express the function under investigation, instead of focusing on a few pairs of seemingly equivalent markers in the two languages.

A first central objective of this book is to revisit the largely introspection-based claims available in the contrastive literature by comparing the frequency and preferred grammatical types (viz. conjunctive adjuncts, coordinators and subordinators) of conjunctive markers used by each language to express contrast between two discourse segments, using large corpora of authentic data in the two languages. In addition, this research also pays special attention to the syntactic features of English and French conjunctive markers of contrast. The syntactic analysis of the markers revolves around two main axes. Firstly, the study investigates the types of syntactic structures in which English and French conjunctive markers of contrast tend to be included. In current discourse research, it is generally accepted that the basic unit linked by conjunctive markers corresponds to the clause. Markers that link smaller units than the clause, by contrast, do not perform a conjunctive (or discourse) function (e.g. Mann & Thompson 1988; Miltsakaki et al. 2004; Hoek et al. 2018). In both English and French, the clauses hosting conjunctive markers may take a variety of forms: for example, they may either stand alone in discourse – as in (3), or be syntactically dependent on another clause – as in (4) and (5); they may be built around a finite or a non-finite verb form (as in (3) and (4), respectively), or even take the form of a verbless segment, as in (5).

(3)   **However**, that does not exonerate Mr Blair from blame for the slipshod processes and slapdash presentation over which he has presided.

(4)   They appear to be winning, **although** not without inflicting serious casualties on the civilian population.

(5)   Hopes of a wake-up call from the electorate, **while** not impossible, look slim.

The types of syntactic structures in which conjunctive markers typically occur have been granted very little attention in the research to date, which has focused heavily on the semantic features of these linguistic units. Yet, as this study will attempt to demonstrate, the choice of one type of syntactic structure over the other(s) for a given conjunctive marker is not without significance but can be indicative of the strategies of textual development adopted by the writer. More specifically, the

analysis of corpus data will show that certain syntactic uses of conjunctive markers may create a range of stylistic effects at discourse level, such as emphasis or compression. Accordingly, a comparison of the syntactic patterning of English and French conjunctive markers can shed light on subtle differences in the discourse strategies employed by each language with respect to the expression of contrast.

The second syntactic feature of conjunctive markers investigated in this book is their position in the sentence. This part of the study zooms in on the only category of markers which are syntactically mobile, viz. conjunctive adjuncts (e.g. *however, cependant*). Unlike the syntactic features of the segments hosting conjunctive markers, the placement patterns of conjunctive adjuncts have been at the centre of a number of corpus-based studies, including contrastive ones: some researchers have demonstrated that languages differ in the positions that they tend to prefer for conjunctive adjuncts, even when they offer a similar range of placement possibilities for these items (cf. e.g. Altenberg 2006 on English and Swedish; Kunz & Lapshinova-Koltunski 2014 on English and German; Balažic Bulc & Gorjanc 2015 on Croatian and Slovene). However, such a comparison remains to be carried out for the English-French language pair. In addition, the research currently available – whether mono- or multilingual – tends to make fairly general observations on the placement tendencies of the languages analysed (and/ or the cross-linguistic differences between them), without considering the possibility that these may be influenced by the communicative situation. The present study aims to compare the placement patterns of English and French conjunctive adjuncts of contrast in a large corpus of authentic texts. In line with the recent insistence, in corpus-based contrastive linguistics, on the importance of taking register variation into account when formulating differences between languages (e.g. Lefer & Vogeleer 2014), the placement patterns of English and French conjunctive adjuncts are investigated in two distinct registers (viz. academic prose and newspaper editorials). The objective is to assess the impact of register variation not only on the placement tendencies of each language taken individually, but also on the cross-linguistic differences between them.

In addition, both parts of the syntactic analysis will examine the influence of lexis on the patterns uncovered within each language system. This objective derives from the keen awareness, in current text-based linguistics, that the grammatical (including the syntactic) patterns observed in texts are inextricably linked to the lexical choices made by writers and speakers (cf. e.g. Sinclair 1991; Hunston & Francis 2000; Hoey 2005). Amongst other aspects, this implies that even when they belong to the same grammatical category, different lexical items may display partly diverging, idiosyncratic patterns of use, be it in terms of the types of syntactic structures in which they are typically found, the (type of) linguistic elements with which they tend to co-occur, their position in the sentence, etc.

In other words, while they are characterised by a certain degree of regularity and repetitiveness, the usage patterns observed in a given language system are also partly item-specific, and thus can be neither predicted nor captured by fully general rules (see Herbst et al. 2014: 1–2). As a result, in addition to carrying out fairly general comparisons of the syntactic patterns of use of conjunctive markers of contrast across languages and registers taken as wholes, the present research will also account for the role played by lexical variation on both (i) the types of clauses hosting conjunctive markers, and (ii) the positional preferences of English and French conjunctive adjuncts of contrast. Finally, whenever possible, this study will also look for possible interactions between the various factors of influence (viz. language, register and lexis) included in the analysis (for example, by assessing whether the amount of lexical variation uncovered in the data varies in function of the language or register analysed).

To sum up, the present book aims to answer two main research questions:

1. *Frequency and patterns of use of English and French conjunctive markers of contrast*
    Do English and French differ in (i) their overall frequency of use of conjunctive markers of contrast; and (ii) the types of conjunctive markers that they tend to prefer to signal contrast between two clauses or sentences?

2. *Placement patterns of English and French conjunctive adjuncts of contrast*
    Do English and French differ with respect to the positions that they allow for and/or prefer for conjunctive adjuncts of contrast?

With respect to the first of these research questions, two hypotheses may be formulated on the basis of the non-empirical English-French contrastive literature. Firstly, in terms of the overall frequency of explicit conjunctive markers, French is expected to display significantly more markers of contrast than English. Secondly, regarding the preferred types of markers used in each language, it is hypothesised that logico-semantic relations of contrast will tend to be signalled more frequently by means of coordinators in English, and subordinators as well as conjunctive adjuncts in French. In addition, this part of the study will also include more detailed comparisons of conjunctive marker use in the two languages. Rather than simply comparing the frequency of conjunctive adjuncts, coordinators and subordinators of contrast in English and French, I will also investigate the types of syntactic structures in which these items tend to be involved in the two languages.

In the second part of the book, English and French are compared with respect to both (i) their possible and (ii) their preferred syntactic positions for conjunctive adjuncts of contrast. In the monolingual literature, English conjunctive adjuncts have largely been associated with the sentence-initial position (e.g. Quirk et al.

1985: 643; Leech & Svartvik 2002: 139; Halliday & Matthiessen 2004: 83). In French grammars, on the other hand, emphasis has also been laid on the tendency of conjunctive adjuncts to occur sentence-medially, and more particularly within the verb phrase (e.g. Grevisse & Goosse 2011: 1211) – although such statements largely remain to be verified empirically. Therefore, it can be hypothesised that French will display a greater proportion of conjunctive adjuncts used sentence-medially than English, where a majority of conjunctive adjuncts should appear at the beginning of the sentence. In addition to identifying general differences between English and French patterns of conjunctive adjunct placement, the analyses carried out in this study will help to assess the impact of both register and lexis on conjunctive adjunct placement in the two languages, and to identify possible interactions between these factors with respect to placement.

1.3   Framework of the study

The present study adopts a combined Systemic Functional and corpus linguistic approach to the study of conjunctive markers of contrast. As has been demonstrated by a number of leading Systemic Functional linguists in the past fifteen years, many benefits can be gained from the disciplinary synergy between Systemic Functional theory and corpus methodology (e.g. Butler 2004; Thompson & Hunston 2006a; 2006b; Halliday 2006; Matthiessen 2006; Hunston 2013). On the one hand, the two linguistic approaches share a number of central concerns – such as the importance attached to the analysis of authentic data or the focus laid on register variation – which create a "natural affinity" between them, to use Halliday's (2006: 293) own words. In fact, as made clear by Thompson and Hunston (2006b), the two frameworks are really complementary: as a predominantly theory-oriented approach to language, Systemic Functional Linguistics can increase the degree of theoretical elaboration that has sometimes been said to lack in corpus work. More precisely, Systemic Functional Linguistics makes it possible to bring order to some of the findings emerging from corpus analyses, in response to the criticism that corpus linguistics too often gives rise to "[d]etailed descriptions of individual 'bits of language' [that are] difficult to relate […] to an overall picture of language" (Thompson & Hunston 2006b: 3). Conversely, the solid empirical basis, along with the powerful methods of extraction and analysis, afforded by corpus linguistics can considerably expand the scope – and hence the generalisability – of the Systemic Functional theoretical descriptions, which have so far remained essentially based on rather limited amounts of authentic data analysed fully manually.

　　　　Yet, despite the recent insistence on the many advantages that can be gained from the combination of Systemic Functional theory with corpus methodology,

little research has yet exploited this disciplinary synergy to the full, by truly apply-
ing the main tenets of both approaches simultaneously (but see e.g. Neumann
2013; Kunz & Lapshinova-Koltunski 2014; Zinn & McDonald 2018 for some nota-
ble exceptions). The present book situates itself in a set of recent studies which
have worked towards a greater integration of the Systemic Functional and corpus
linguistic frameworks, with a view to making both approaches benefit from what
the other has to offer. In addition to the objectives described in the previous sec-
tion, the study thus also pursues a more theoretical goal. On the one hand, I will
test the degree of descriptive accuracy and generalisability of the aspects of the
Systemic Functional theory under investigation here, by confronting them to large
bodies of authentic data across languages and registers. If it proves necessary, these
theoretical aspects will be refined in order to better reflect the range of functions
and meanings observed in authentic linguistic productions. Conversely, great care
will be taken to relate the tendencies emerging from the corpus data to Systemic
Functional concepts and categories, so as to avoid the kind of disparate, theory-
thin descriptions for which corpus linguistics has sometimes been criticised (cf.
Thompson & Hunston 2006b: 3–4; Hunston 2013: 618).

The corpus data used in this study consists of comparable original written texts
in English and French representing two distinct registers – making it possible to
integrate register-sensitive considerations into (some of) the cross-linguistic com-
parisons. The first subcorpus is the *Multilingual Corpus of Editorials* (Mult-Ed),
and contains editorials from a range of quality newspapers in the two languages.
The second subcorpus, viz. the *Louvain Corpus of Research Articles* (LOCRA), is
made up of research articles from five disciplines in the Humanities (viz. anthro-
pology, education, political science, psychology and sociology). Each subcorpus
contains about 2 million words of original texts per language, amounting to about
8 million words in the entire corpus.

1.4   Outline of the book

This book is structured in seven chapters, grouped in three main subparts. The
first part of the book provides the necessary theoretical background for the study
and consists of three chapters. In Chapter 2, the two theoretical constructs that
are central to this research, viz. conjunctive markers and contrast, are defined.
Throughout the chapter, special attention is granted to issues of cross-linguistic
comparability, by constantly making sure that the definitions of these two con-
cepts are equally adapted to the description of English and French. Chapter 3 then
provides a general overview of the corpus-based contrastive research on conjunc-
tive markers. After a brief presentation of the field of corpus-based contrastive

linguistics, the chapter provides a state-of-the-art review of the types of corpus-based contrastive approaches to conjunctive markers currently available in the literature. A final part of the chapter zooms in on some cross-linguistic research that focuses more specifically on the English-French language pair. Finally, Chapter 4 is devoted to the relationship between the Systemic Functional and the corpus linguistic frameworks, which are both central to this study. It starts with a discussion of the benefits of such a combined approach, then focuses more specifically on one aspect on the Systemic Functional theory that I intend to explore through the application of corpus methods, viz. the textual metafunction.

The second part of the book is concerned with issues of methodology. Chapter 5 offers a broad presentation of the corpus data and outlines the main methodological steps taken in the study. It also describes the statistical procedures followed in the corpus analyses. Chapter 6 expands on one particularly important stage of the methodology, viz. disambiguation. Disambiguation is the process whereby automatically-extracted instances of the potential conjunctive markers in the corpus are examined in context, so as to identify the occurrences which correspond to the definition of conjunctive markers of contrast adopted in this study, while weeding out those that do not. The disambiguation process consists of two main stages, viz. semantic disambiguation – which aims at discarding the occurrences that do not express a contrastive meaning – and syntactic segmentation – whose objective is to weed out tokens that operate below the level of the clause, and thus are not considered to perform a discourse function.

The third and final part of the book presents the results of the corpus analyses and is divided into two chapters. Chapter 7 compares the frequency and patterns of use of English and French conjunctive markers of contrast. It is the part of the study in which the claim that French tends to use more explicit conjunctive markers than English, together with the assertions concerning the preferred types of markers used in the two languages (viz. coordinators in English, subordinators and conjunctive adjuncts in French), are tested on a large empirical foundation. In a second stage, Chapter 7 also analyses the type of syntactic structures in which English and French conjunctive markers of contrast are used. Chapter 8 then investigates the placement patterns of English and French conjunctive adjuncts of contrast, while also assessing the impact of both register and lexis on the phenomenon. The book ends with a general conclusion summarising the main findings, taking stock of the major contributions made by the study, and suggesting promising avenues for further research.

CHAPTER 2

Defining the key constructs

In this chapter, the two main constructs that are central to this research, viz. conjunctive markers and contrast, are defined. As the present study adopts a contrastive perspective, one of the main concerns here will be to make sure that these definitions are equally valid for English and French. Issues of comparability lying at the core of the chapter, it seems useful to start with a brief discussion of some of the main principles necessary to ensure a reliable cross-linguistic comparison. This is what Section 2.1 will be devoted to. Section 2.2 will then define conjunctive markers, and Section 2.3 will be dedicated to the semantic relation of contrast.

2.1 Issues of comparability for contrastive analysis

One of the most crucial – and probably thorniest – issues in contrastive analysis relates to the notion of equivalence and consists in determining what units to compare across languages. In order to ensure the reliability of a cross-linguistic comparison, it is essential for the contrastivist to make sure that s/he is comparing units that are in fact comparable. This means that before carrying out any cross-linguistic analysis per se, researchers in contrastive linguistics must identify a common, unbiased basis for comparison against which the similarities and differences between languages can then be identified. Although the question of equivalence was given primary importance in the work of early contrastivists (see James 1980; Krzeszowski 1990; Chesterman 1998), many of the more recent studies in contrastive linguistics do not mention them explicitly (Lefer & Cartoni 2011: 90) and tend to take for granted the comparability of their objects of study (Krzeszowski 1990: 36). This book intends to place issues of comparability at the centre of its concerns. In this perspective, the present section provides an overview of the main theoretical reflections pertaining to the question of equivalence for contrastive analysis.

2.1.1 Equivalence and tertium comparationis

At the heart of any reliable cross-linguistic comparison lies the question of the comparability of the phenomena investigated across languages. As made clear by

James (1980: 169) in his seminal book on contrastive analysis, when setting about comparing two languages, "the first thing we do is make sure that we are comparing like with like". In order to comply with this comparability requirement, all contrastive studies must start with the assumption that the compared objects have something in common: as argued by James (ibid.), "the two (or more) entities to be compared, while differing in some respects, must share certain attributes […] since it is only against a background of similarity that differences are significant" (see also Krzeszowski 1990: 15). This "background of similarity" is referred to in contrastive linguistics as the *tertium comparationis*. More precisely, the *tertium comparationis* can be defined as a common, language-neutral platform of comparison used as reference, and against which similarities and differences between languages can be identified (Lefer & Cartoni 2011: 96). Thanks to the *tertium comparationis* (henceforth: TC), the languages are not compared directly to each other. Instead, they are compared through the lens of some third, language-neutral element, so as to make sure that the comparison is not biased in favour of one of the languages – in line with Chesterman's (1998: 29) statement that "no comparison can be made between two entities without a frame of reference provided by a third term of some kind" (see also Pekelder 2010 on the importance of a third term of comparison; Haspelmath 2010 on the necessity of language-neutral categories of comparison). It is noteworthy that, as made clear by several theoreticians in contrastive linguistics, the kind of equivalence required by the *tertium comparationis* is relative: rather than absolute identity, what is needed is maximum similarity between the units compared (see James 1980: 168; Chesterman 1998: 37; Connor & Moreno 2005).[1]

The type of equivalence by virtue of which units are compared across languages can be of various natures, and in addition to the absolute necessity of defining a TC at all, it is also essential to select the appropriate TC for the specific research goals of a given study. The type of TC chosen can have a tremendous impact on the outcome of the comparison: as Krzeszowski (1990: 15) explains, "[d]epending on the platform of reference (or *tertium comparationis*) which we adopt, the same objects turn out to be either similar or different". Krzeszowski (ibid.) illustrates this point by means of a telling example borrowed from geometry:

---

1.   The term equivalence is sometimes used interchangeably or in very similar contexts as *tertium comparationis*. These concepts, while not completely synonymous, stand in a very close relationship. As explained by Krzeszowski (1990: 21), they are two sides of the same coin: equivalence is the principle whereby some TC can be established, as only equivalent items are comparable across languages; on the other hand, the extent to which a TC can be found for two items across languages determines the degree to which these items are equivalent (see also Chesterman 1998: 29).

> [A]ny two or more objects can be compared with respect to various features and, as a result, the compared objects may turn out to be similar in some respects but different in others. Thus, a square and a rectangle are similar in that both consist of four sides at right angles. But they are also different, since in a square, but not in a rectangle, the four sides are of equal length. If we compare squares and rectangles with respect to the angles, we ascertain that the two types of figures are identical. If, on the other hand, we consider the length of their sides, we find them to be different.

Another example, in linguistics this time, is provided by Jaszczolt (2003: 442), who discusses the concept of *tertium comparationis* in relation to semantics and pragmatics. She contrasts the following two sentences in English and Polish:

(1)   Thank you.

(2)   To tylko stara sukienka. [It's only an old dress.]

As she explains, while semantically, syntactically and lexically there are hardly any similarities between these two sentences, pragmatically they are equivalent to some extent, in that they may both function as answers to a speech act of complimenting in their respective language: when complimented, English speakers will thank their interlocutor, whereas Polish speakers will tend to produce a self-deprecating response. As the outcome of the cross-linguistic comparisons greatly depends on the type of TC selected, the choice of the appropriate TC is a crucial component of contrastive analysis.

2.1.2   Various types of equivalence

It is generally accepted among contrastivists that the type of TC chosen fundamentally depends on the object of study: different kinds of analysis (e.g. phonological vs lexical vs pragmatic, etc.) require different types of basis for comparison or TC (see e.g. Chesterman 1998: 29; Gast 2012). However, the question of what type of TC is best-suited for what type of contrastive analysis has aroused much controversy among contrastive linguists. Worse still, in many contrastive studies, decisions concerning the type of basis for comparison chosen for the analysis are simply left implicit. This all ties in well with Krzeszowski's (1990: 15) claim that "the most fundamental concept [in contrastive linguistics] remains as hazy as ever". Traditionally, theoretical discussions in contrastive linguistics have mostly been restricted to two main types of TC, namely formal correspondence – based on the comparison of corresponding grammatical systems or categories across languages – and semantic equivalence – which relies on the identity of meaning between the units compared (see Krzeszowski 1984; Krzeszowski 1990: 16; Chesterman 1998: 30). However, as Krzeszowski (1990: 16) points out, these two

types of equivalence are not a sufficient basis to account for cross-linguistic differences at all linguistic levels: for phonological, pragmatic and discourse studies, for instance, other types of equivalence will have to be found. In addition, each type comes with its own range of limitations and difficulties.

On the one hand, a comparison solely based on a formal *tertium comparationis* is "at best […] incomplete, at worst it cannot be performed at all, and in many cases it is misleading" (Krzeszowski 1990: 16). One obvious reason for this is that such analyses are only possible when the two languages have grammatical categories in common, and that even when they do, grammatical labels may be the same across languages but not refer to exactly the same phenomena (James 1980: 169–170; see also Haspelmath 2010 on the lack of comparability of grammatical categories across languages). To illustrate the flaws of a purely formal TC, Krzeszowski (1990: 16) gives the example of a comparison between the present perfect in English and the *passé composé* in French: in this case, the formal similarity between the two constructions (i.e. auxiliary + past participle) is not matched by a semantic similarity, since these two tenses do not systematically appear in the same situations – as attested by the considerable difficulties that the present perfect poses for French-speaking learners of English, for example. Another problem is that different languages do not necessarily use the same grammatical devices to express similar things. For example, Johansson (2007: 3) notes that the meaning expressed by modal auxiliaries in one language may be expressed by different linguistic devices in another language (e.g. adverbs such as *probably, perhaps*), so that a comparison based on modal auxiliaries only would be unreliable. Therefore, it is generally accepted that a formal TC will necessarily require support from other types of equivalence, for example semantic or functional equivalence. In Section 2.2, we will see that the limitations inherent in a formal TC are especially relevant for the study of conjunctive markers, which constitute a functional category made up of elements from various grammatical classes.

A *tertium comparationis* based on semantic equivalence is generally accepted to be preferable to a formal one. Nevertheless, it is not free of difficulties either. The main issue with semantic equivalence is that it is generally equated with translation equivalence. Yet, as Krzeszowski (1990: 17) points out, "translation equivalents are often semantically non-equivalent". Translations may deviate from semantic equivalence for a number of reasons, including: (i) translation errors; (ii) diverging formal properties of the languages involved in the comparison; (iii) stylistic reasons; (iv) translation-induced effects, such as source-language influence or translation universals; and (v) creative translations and free renderings (Krzeszowski 1990: 17; Altenberg & Granger 2002: 17). As a result, when used as a TC, translation equivalence should be approached with caution. Several solutions are available to better control for the degree of equivalence of translations,

such as the recourse to Ivir's (1983) procedure of back-translation, the reliance on recurrent translational patterns – in line with Krzeszowski's quantitative notion of statistical equivalence – or the use of Altenberg's measure of mutual correspondence, which calculates how frequently two items are translated by each other (see Altenberg & Granger 2002 for an overview). Provided that it is rigorously established, a semantic *tertium comparationis* can yield fruitful results. An illustration of an efficient semantic TC is found in Lefer and Cartoni's (2011) methodological article on the comparison of derivational prefixes in English, French and Italian. After establishing a definition of the morphological category of derivational prefixes and collecting as many prefixes corresponding to that definition as possible in the three languages, Lefer and Cartoni set up a semantic TC made up of six major semantic categories (viz. location, evaluation, negation, quantity, modality and inchoativity). They classify each English, French and Italian prefix into one of these six categories (or several of them, for polysemous prefixes), which gives them a general picture of the number of prefixes available in each subcategory in each language. This makes it possible to compare and contrast the resources available in each language in a consistent way, while "avoiding basing [their] contrastive study on a (more or less) arbitrarily chosen set of prefixes" (ibid.: 99). To give a more concrete example, based on their TC, Lefer and Cartoni have identified that reversal and removal are expressed by the following set of prefixes: (i) *de-, dis-, un-* in English; (ii) *dé-, dis-* in French; and (iii) *de-, s-, dis-* in Italian (ibid.). Thanks to this semantically-based inventory, in order to compare the frequency of use of prefixes expressing reversal and removal across the three languages, the researchers would merely need to 'count' the number of each of these prefixes in a comparable corpus of French, Italian and English (which would still necessitate disambiguation in the case of polysemous prefixes), and compare the frequencies. In this way, "[r]ather than comparing the productivity of individual prefixes across languages (which may prove to be problematic, for example, when a given affix in language A has no equivalent in language B), the TC makes it possible to contrast semantic groupings of prefixes, thereby allowing interesting generalizations to be made" (ibid.: 95).

Although a semantic TC seems preferable to a formal one in some situations, as mentioned earlier this type of equivalence is not appropriate to carry out cross-linguistic comparisons at all linguistic levels. As a result, some researchers have suggested alternative types of TC for the various levels of linguistic analysis. Krzeszowski (1984; 1990: chap. 3), for instance, still unconvinced of the merits of either a formal or a semantic TC used on their own, establishes a typology of seven types of TC (namely statistical equivalence, translation equivalence, system equivalence, semanto-syntactic equivalence, rule equivalence, pragmatic (functional) equivalence, substantial equivalence) which can be selected and combined depending

on the type of comparison carried out. James (1980), on the other hand, identifies three main types of TC, each emerging as the best available equivalence for one sort of linguistic comparison: (i) the International Phonetic Alphabet and the vowel diagrams for phonological contrastive analysis; (ii) a set of supposedly universal semantic components for cross-linguistic lexical studies; and (iii) translation equivalence for cross-linguistic comparisons of grammar (see James 1980: 166–178; see also Willems et al. 2004 for a discussion of the various types of TC to be used at different levels of linguistic analysis). For higher-level linguistic phenomena, the prevailing basis for comparison seems to be pragmatic (also called 'functional') equivalence. As stated by Chesterman (1998: 35), pragmatic equivalence "is the only [type of equivalence] that is explicitly said to pertain to texts". In its most broadly accepted sense, however, pragmatic equivalence does not seem to be suitable for all types of macrolinguistic analyses and has predominantly been associated with equivalence in terms of speech acts. Thus, Olesky (1984: 360), for example, defines pragmatic equivalence as follows: "[a] linguistic expression X_1L_1 is pragmatically equivalent to a linguistic expression X_2L_2 if both X_1 and X_2 can be used to perform the same speech act […] in L_1 and L_2" (see also Janicki 1990; Jaszczolt 2003 for more on pragmatic equivalence). Alternative accounts have nevertheless relied on functional equivalence in particular (i.e. as opposed to pragmatic equivalence) for the analysis of other macrolinguistic phenomena, such as cohesion for example (Moreno 1998; Kunz & Steiner 2013; see also Chesterman's 1998 'contrastive functional rhetoric'). In these studies, the grounds for comparing units across languages is that they perform the same function in text (e.g. linking two textual units). In summary, to this day it is still somewhat unclear exactly what kind of TC must be used for what type of contrastive study. As Lefer and Cartoni (2011: 97) rightly put it, whatever the type of *tertium comparationis* selected, it is in any case crucial for the researcher to state his/her decision explicitly.

So far, the *tertium comparationis* has mainly been discussed in relation to the linguistic units to be analysed cross-linguistically. However, it is worth mentioning that Connor and Moreno (2005) advocate a broader approach to the TC, which consists in finding a common basis for comparison at all levels of the analysis, including the data selection process. Thus, for them, the issue of data comparability, which supposes that the texts included in the corpora must be comparable in terms of genre, mode, type of participants, etc., is also a type of *tertium comparationis*. The article in fact expands on an idea put forward by Krzeszowski (1990: 26), who argued that each contrastive study is characterised by several subtypes of *tertium comparationis*, each being concerned with one type of equivalence. One such subtype is meant to ensure the comparability of the texts included in the study. In the present research, issues of data comparability are not considered to be within the purview of the TC. While the concept of *tertium comparationis* is

purely specific to contrastive linguistics, the necessity to ensure the comparability of the data is rather seen as a general preoccupation that is common to all types of comparative endeavours in linguistics (e.g. diachronic studies, variationist socio-linguistics, learner corpus research, etc.). The crucial issue of data comparability will therefore be tackled in Chapter 5.

2.1.3   Issues of circularity in contrastive linguistics: Chesterman's Contrastive Functional Analysis

One final issue of which the contrastive researcher must be aware when defining the object of his/her cross-linguistic comparison pertains to the risk of circularity. Once a reliable *tertium comparationis* has been established, the contrastivist can set about comparing the relevant linguistic phenomena across languages. This comparison stage can give rise to three possible outcomes (see Krzeszowski 1990: 37–38): it can reveal that (i) a given item X in language A is identical in some respects to an equivalent item Y in language B; (ii) a given item X in language A is different in some respects from an equivalent item Y in language B; and (iii) a given item X in language A has no equivalent in language B. However, as Chesterman (1998: 52) points out, if equivalence is the justification for comparing two items, then how can it also be the result of the comparison (as is the case in the first situation)? In 1990, Krzeszowski (1990: 20) had already raised awareness to the issue of circularity, which he had defined as follows:

> The circularity consists in the following: we compare in order to see what is similar and what is different in the compared materials; we can only compare items which are in some respect similar, but we cannot use similarity as an independent criterion in deciding how to match items for comparison since similarity (or difference) is to result from the comparison and not motivate it.

Against this backdrop, Chesterman (1998) puts forward a methodology designed to avoid circularity in cross-linguistic studies, which he calls Contrastive Functional Analysis (CFA). He suggests a recursive methodology in seven steps, based on a relative view of equivalence. The starting point of a CFA-based study is an assumed or perceived similarity between two linguistic phenomena X and Y across languages, instead of an absolute statement of equivalence.[2] This perceived similarity is the justification for carrying out the cross-linguistic comparison and will give rise to a hypothesis – inspired from the null hypothesis in statistics – which assumes the identity of X and Y across languages, and which the study sets out to reject, or at least refine. The hypothesis is falsifiable and will therefore be

---

2.   Note that this idea was already touched upon by James (1980: 169).

tested empirically. The objective of the testing stage will be to state to what extent the identity hypothesis can be supported or rejected, and on what conditions. In other words, the result of the testing phase is the statement of the ways in which X and Y differ, and of the ways in which they are similar. In Chesterman's CFA methodology, this is where the *tertium comparationis* makes its appearance. Here it is defined quite differently from previous accounts, not as the starting point of the comparison of linguistic items across languages, but as the evidence in favour of the hypothesis. As Chesterman explains (1998: 58):

> In this methodology, the *tertium comparationis* is thus what we aim to arrive at, after rigorous analysis: it crystallizes whatever is (to some extent) common to X and Y. It is thus an explicitation of the initial comparability criterion, but it is not identical with it – hence there is no circularity here. Using an economic metaphor, we could say that the *tertium comparationis* thus arrived at adds value to the initial perception of comparability, in that the analysis has added explicitness, precision, perhaps formalization; it may also have provided added information, added insights, added perception.

Thus, the perception of similarity constituting the starting point of the CFA methodology is revised and refined through a testing process to produce what Chesterman calls the *tertium comparationis*. Based on the results of the testing stage, the researcher formulates a revised hypothesis in which the relationship between X and Y is redefined: X and Y are no longer viewed as identical, but as similar in some respects, and different in others. This revised hypothesis can in turn be tested, in an iterative process aiming to provide the most precise picture possible of the similarities and differences between the two items investigated across languages.

Given the polysemy of the term *tertium comparationis*, a terminological comment is in order here: in this book, *tertium comparationis* is used in its more traditional meaning, as a common basis for the cross-linguistic comparison of linguistic phenomena. However, the present study also adopts the kind of falsificationist approach advocated by Chesterman to avoid circularity: starting from an assumption of similarity, it aims at refining this initial assumption by pinpointing, on the basis of corpus-based analyses, the specific ways in which English and French conjunctive markers of contrast differ, and the ones in which they are similar.

The remainder of the present chapter will be devoted to the two concepts that are central to this study, i.e. conjunctive markers and contrast. While defining these two categories, I will strive to apply the principles described in this section as scrupulously as possible: the purpose will be to arrive at reliable *tertia comparationis* (in James's and Krzeszowski's sense) for my two research constructs, in order to make sure that the subsequent comparisons will not be skewed in favour of either English or French.

2.2   Conjunctive markers

The present section sets out to provide a reliable definition of conjunctive markers for the cross-linguistic comparison of these items in English and French. In accordance with the theoretical framework adopted in this study, the definition of conjunctive markers will rely heavily on the description of cohesion and conjunctive markers provided in Systemic Functional Linguistics. It should be noted that a substantial portion of the research in Systemic Functional Linguistics (henceforth: SFL) has been concerned with the English language, as attested by the fact that Halliday's seminal book *An Introduction to Functional Grammar* (1985) provides a description of English. While the framework has also been successfully applied to French on several occasions (e.g. Banks et al. 2009; Caffarel 2006; Banks 2017), Systemic Functional descriptions of French *cohesion* remain scarce. Therefore, this section relies heavily on literature pertaining to the description of cohesion in English. Nevertheless, as pointed out by Caffarel et al. (2004: 6), although many *descriptions* within the Systemic Functional framework pertain to English, the Systemic Functional *theory* was always meant to be a general one, virtually applicable to the description of all languages. The book edited by Caffarel et al. (2004), which contains Systemic Functional descriptions of languages as diverse as German, Japanese, Vietnamese, Telugu and Pitjantjatjara, provides evidence that the multilingual ambition of Systemic Functional Linguistics is largely achieved, as does the fairly large body of research that has been undertaken to apply Systemic Functional Linguistics to a collection of other languages including Arabic, Dutch, Indonesian, Spanish, Swedish (see e.g. Mwinlaaru & Xuan 2016 for an overview). There are therefore good reasons to believe that descriptions of cohesion in English are largely applicable to the description of the French system as well.

This section is made up of five main parts. Section 2.2.1 provides a definition of the general notion of cohesion, and specifies how it relates to coherence. Section 2.2.2 zooms in on the subtype of cohesion that is the focus of this research, i.e. conjunction. Section 2.2.3 presents the main approaches to conjunction and conjunctive markers that are found in the Systemic Functional literature, and Section 2.2.4 explains why a broad approach to conjunction is adopted here. Finally, Section 2.2.5 presents some core features of conjunctive markers.

2.2.1   Cohesion and coherence

The concept of cohesion was developed by Halliday and Hasan (1976) in their seminal book *Cohesion in English*. It refers to one of the two properties of texts that provide them with texture, i.e. that make a given linguistic stretch emerge as a text, functioning as a unified whole, instead of a collection of unrelated sentences

(ibid.: 1).[3] The system of cohesion offers a range of possibilities for linking lin-guistic units to what has gone before, and has to do with "the way we relate or tie together our bits of discourse" (Eggins 1994: 88) by setting up relations of interpretative dependence between them. More specifically, Halliday and Hasan (1976: 4) define cohesion as follows:

> Cohesion occurs where the INTERPRETATION of some element in the discourse is dependent on that of another. The one PRESUPPOSES the other, in the sense that it cannot be effectively decoded except by recourse to it. When this happens, a relation of cohesion is set up, and the two elements, the presupposing and the presupposed, are thereby at least potentially integrated into a text.

One central property of cohesion as defined by Halliday and Hasan is that it is pri-marily a semantic rather than a structural or grammatical relation. It is intended to account for properties of texts that cannot be explained by purely grammati-cal factors: as Halliday and Hasan (ibid.: 6f) explain, the sentences in a text are not related by any structural means, and yet language users are able to determine whether a given linguistic passage functions as a text or not. Therefore, there must be linguistic properties other than grammatical structure which govern the con-struction of texts (see also Butler 2003: 336). The main purpose of cohesion stud-ies is to uncover (some of) these non-structural, semantic mechanisms.

Five main types of cohesive resources have traditionally been identified, viz. reference, substitution, ellipsis, conjunction and lexical cohesion. Reference occurs when two (or more) expressions are used to refer to the same referent in a text. The cohesion therefore lies in the continuity of referents, with the same entity being brought up several times by means of different linguistic units (Halliday & Hasan 1976: 31). An illustration of reference is provided in (3), where the interpretation of the pronoun *it* depends on its being identified as co-referent with *their bedroom*. Typical realisations of reference include pronouns (*he, she, they*) or determiners (*this, that, those*).

(3)   They came again into their bedroom. A large bed had been left in it.

(Thompson 2014: 216)

Like reference, substitution consists in replacing a linguistic element with another one to avoid repetition, as in (4). The difference between reference and substitu-tion is that substitution is "a relation in the wording rather than in the mean-ing" (Halliday & Hasan 1976: 88). While with reference the cohesive device and

---

3.   The other property is situational coherence, i.e. the property of displaying recognisable and consistent features in terms of register, in the sense that "we can think of one situation in which all the clauses of the text could occur" (Eggins 1994: 87).

its antecedent refer to the same entity (e.g. *their bedroom* and *it* refer to one and the same bedroom), with substitution the two expressions refer to two distinct (although similar) entities (e.g. in this case: the speaker's current axe, and another, sharper one which s/he intends to buy).

(4)   My <u>axe</u> is too blunt. I need a sharper <u>one</u>.   (Halliday & Hasan 1976: 89)

Substitution is very close to ellipsis (and is in fact grouped with it in some accounts, see e.g. Halliday & Matthiessen 2004), which has sometimes been described as 'substitution by nothing'. Ellipsis includes "resources for omitting a clause, or some part of a clause or group, in contexts in which it can be assumed" (Martin 2001: 36). Example (5) displays three instances of ellipsis: *shape* is ellipted twice, and the verb *be* is omitted once.

(5)   In this sketch, there are four shapes. One [ø] is a square, one [ø] [ø]
       a triangle, and there are two circles.   (Christiansen 2011: 20)

The fourth type of cohesive device, conjunction, "refers to how the writer creates and expresses logical relationships [such as causality or contrast] between the parts of a text" (Eggins 1994: 105). The passage below displays two instances of cohesive conjunction, i.e. *for example* and *by contrast*.

(6)   It is easy to identify theoretical conflicts in management accounting.
       For example, contingency theorists argue that the type of management
       accounting system which is appropriate to an organization is dependent
       on a number of organization-specific variables. **By contrast**, the emphasis
       in much of the management accounting research published between
       the late 1950s and the mid-1970s was on the development of specific
       normative models which were allegedly suitable for use in a wide variety
       of organizations without any context-specific adaptation.
                                                              (Bloor & Bloor 2004: 98)

Finally, lexical cohesion is "the cohesive effect achieved by the selection of vocabulary" (Halliday & Hasan 1976: 274). It encompasses a wide array of lexical phenomena, such as the repetition of a given word several times through a text, the recourse to (near-)synonyms, general and superordinate terms, collocations, and more generally words belonging to the same semantic field (see e.g. Hoey 1991 for more details). Example (7) contains several instances of lexical cohesion: for example, *flower, stigmas* and *pollen* are repeated; in addition, the words related to *pollen* (*pollination, pollinate*) also contribute to the lexical cohesion of the passage.

(7)   A flower cannot produce seeds until it is pollinated and its ovule fertilized.
       Pollination is the transfer of pollen from the male parts (stamens) to the
       female parts (stigmas) of a flower. If pollen is carried to the stigma of the
       same flower, it is called self-pollination.   (Bloor & Bloor 2004: 99)

Ever since it was developed by Halliday and Hasan, cohesion has triggered a plethora of research and has been approached from a variety of perspectives. Some of its aspects have generated debate, including the sharp dividing line drawn between grammatical and non-grammatical resources (an issue that will be taken up in Section 2.2.3), or the number and types of cohesive resources that exist and the ways in which to classify them (see Xi 2010 for an overview). For example, while a majority of studies on cohesion have largely built on Halliday and Hasan's work, de Beaugrande and Dressler (1981) identify not five but ten cohesive categories, adding (among others) tense and aspect to the set of cohesive resources and using different labels to refer to similar phenomena. Another subject of intense and continuing controversy, which is worth mentioning here, is the relationship between cohesion and coherence. In Halliday and Hasan's (1976) framework, cohesion is described as one particular kind of coherence: according to them texts can be coherent in two respects: (i) with respect to the context of situation in which they occur, and therefore be coherent in terms of register; and (ii) with respect to themselves, and therefore be cohesive. For them, the two concepts are thus inseparable and interdependent, and both types of coherence are necessary to define a text (ibid.: 23). This view of coherence has however been severely criticised for its insistence on the *necessity* of cohesion for the elaboration of texts. Many researchers have tried to demonstrate that cohesion is secondary in comparison to coherence: on the one hand, they argue, a text containing a high number of cohesive devices can very well completely lack coherence, as shown in Enkvist's oft-cited example in (8) (where all the instances of lexical cohesion are italicised); on the other hand, a text with no cohesive markers can be coherent, as in (9) (see also Tanskanen 2006: 16f; Christiansen 2011).[4]

(8)   I bought a *Ford*. A *car* in which *President* Wilson rode down the Champs Elysees was *black*. *Black* English has been widely *discussed*. The *discussions* between the *presidents* ended last *week*. A *week* has seven *days*. Every *day* I feed my *cat*. *Cats* have four legs. The *cat* is on the *mat*. *Mat* has three letters.   (Enkvist 1978: 110)

(9)   A: That's the telephone
       B: I'm in the bath
       A: OK.   (Widdowson 1978: 29)

In the same vein, a number of researchers have attempted to establish which resource, of cohesion and coherence, takes precedence over the other (i.e. is coherence the

---

4.   Although in practice coherent texts displaying no instances of cohesion are hard to find (Tanskanen 2006: 16).

result of cohesion, or does coherence trigger cohesion?). They have demonstrated that no causal relation seems to exist between the number of cohesive devices in a text and its degree of coherence, thus concluding that coherence cannot be said to be a product of cohesion. In fact, if anything, for these researchers cohesion itself arises from textual coherence, with markers of cohesion acting as signals of relations of coherence that precede them (see e.g. Tierney & Mosenthal 1983; Crewe 1990: 320). Thus, they also establish cohesion as secondary to coherence, and demonstrate that it is not a sufficient condition to achieve coherence. Yet another view, which seems to be the one that prevails today, is that both cohesion and coherence contribute to text creation in a meaningful way, and simply display a difference in nature. While cohesion is a formal property of texts and refers to the surface, linguistic devices that are used to signal relationships between stretches of discourse, coherence is not strictly speaking a linguistic property, but rather pertains to the interpretation or evaluation that the reader/listener makes of a text. Thus, whereas cohesion is an objective property of texts that can be observed and quantified, coherence is more subjective in nature, and a given text may appear coherent to some readers but not to others, depending on their respective world knowledge (see e.g. Hasan 1984; Hoey 1991: 11; Tanskanen 2006: 20–21; Thompson 2014: 215). The two phenomena are both distinct and closely related, in that the presence of cohesive devices facilitates the identification of coherence by the receiver.

2.2.2   A focus on conjunctive cohesion

This research focuses on one type of cohesive device in particular, viz. conjunction. As explained above, conjunction is the cohesive resource through which the logical relationships between textual segments (e.g. contrast, causality, addition) can be signalled explicitly, by means of such devices as *but, in addition, therefore,* etc. Cohesive relations of conjunction are characterised by a number of features that establish them as a separate category with respect to the other types of cohesive resources (see e.g. Kunz & Lapshinova-Koltunski 2014 for an overview). The first specific feature of conjunction has to do with the type of semantic relation that it establishes. Unlike the other types of cohesion, conjunctive devices are primarily relational rather than referring expressions (Kunz & Lapshinova-Koltunski 2014: 238). In relations of reference, substitution or ellipsis, the appearance of a given cohesive tie instructs the reader to look for another element in the text that will allow him/her to interpret it (e.g. in Example (3) above, the reader had to work out that *it* referred to *their bedroom* to be able to interpret the pronoun properly). Conjunctive cohesion, on the other hand, does not convey such a search instruction, but instead provides "a specification of the way in which what is to follow is systematically connected to what has gone before" (Halliday & Hasan 1976: 227).

In other words, as Halliday and Hasan (1976: 236) put it, unlike the other types of cohesion, conjunction is not an overtly anaphoric relation:

> Conjunctive elements are cohesive not in themselves, but indirectly, by virtue of their specific meanings; they are not primarily devices for reaching out to the preceding (or following) text, but they express certain meanings which presuppose the presence of other components in discourse.

Whilst reference, substitution, ellipsis and lexical cohesion are ways of repeating some elements of the message several times through the text, conjunction does not primarily involve such repetitions (Hoey 1991: 6).[5] Another defining feature of conjunction has to do with the syntactic status of the elements connected. While with reference, substitution, ellipsis and lexical cohesion alike, the cohesive devices are (part of) constituents in the sentence in which they occur, conjunctive elements are really in-between elements, as made clear by Christiansen (2011: 161) when he equates conjunctive devices to the mortar that creates a textual wall out of sentential bricks. As a result, strictly speaking, in each conjunctive relation there are three elements, i.e. the two related segments and the cohesive device itself, which points both backward and forward. For other types of cohesion, on the other hand, the related elements stand in a more direct relationship (e.g. in reference, a pronoun refers directly to the noun it stands for, without the presence of a third term ensuring the relationship between them), and each instance of cohesive tie involves only two elements standing *either* in an anaphoric *or* a cataphoric relationship (Kunz & Lapshinova-Koltunski 2014: 237). A final distinctive feature of conjunction pertains to the type of semantic elements that it relates. As explained by Castagnoli (2009: 49ff), while the other types of cohesion may link referents as well as processes, conjunctive cohesion may only link processes. As a result, according to Kunz and Lapshinova-Koltunski (2014: 237), the elements linked by conjunction can be said to display a generally high level of conceptual complexity, which is in turn reflected linguistically by the structural complexity of the elements linked, since conjunction may relate clauses, clause complexes or even entire paragraphs. The other cohesive devices, on the other hand, typically link elements with a lower degree of complexity. For all these reasons, conjunction can be said to "stand on its own as a category [of cohesion]" (Hoey 1991: 5).

Ever since conjunction was first theorised by Halliday and Hasan in 1976, it has triggered an impressive body of research, both inside and outside the realm of SFL. Throughout the years, a large number of accounts on conjunction have

---

5.   Note that some instances of conjunction involving both anaphora and repetition are attested, e.g. *in spite of this, for this reason,* etc. However, neither repetition nor anaphora are defining features of conjunction.

been published, with diverging and sometimes conflicting views of the phenomenon. One largely debated issue has pertained to the number and types of conjunction that may be identified. Initially, four main types of conjunctive relations were identified by Halliday and Hasan (1976), viz. (i) additive (e.g. *and, in addition*, etc. but also some relations of dissimilarity such as *on the other hand, by contrast*); (ii) adversative (e.g. *but, however, instead*); (iii) causal (e.g. *so, as a result*); and (iv) temporal (e.g. *then, meanwhile*). Since then, this classification has been reshaped in a number of ways, so that to this day the number, types, categorisations and denominations of conjunctive relations vary considerably in the literature (see e.g. Martin 1992: 170–178). Halliday himself, for example, redesigned the classification less than ten years later in the first edition of his *Introduction to Functional Grammar* (1985), by redistributing the initial (though sometimes reworked) categories across the three more general classes of (i) elaboration (one element elaborates on the meaning of another by further specifying or describing it); (ii) extension (one unit extends the meaning of another by adding, replacing or alternating some information) and (iii) enhancement (one element enhances the meaning of another by qualifying it by reference to time, place, cause, condition, etc.). The debate pertaining to the semantic categorisation of conjunctive relations, and more particularly relations of contrast, will be further discussed in Section 2.3. Another major area of disagreement among the research community on cohesive conjunction has to do with the level at which it operates. While some researchers, following Halliday and Hasan (1976), consider conjunction to occur exclusively between sentences or larger units, others take relations at lower levels (i.e. between clauses, but also sometimes between smaller units) to be relevant instances of conjunction as well. The stance adopted in this respect naturally has a tremendous impact on the type of linguistic signals investigated (e.g. inclusion or exclusion of such devices as *although, since*, etc., which cannot link units larger than the clause).

2.2.3   Different views on conjunction: Broad and narrow approaches

One of the most vigorously debated issues within the field of research on cohesion concerns the type of linguistic units which cohesive conjunction can be said to relate (see e.g. Martin 1992: 17–21; Butler 2003: 337; Xi 2010; see also Touratier 2006 for a non-SFL account on French). In this respect, two main approaches to conjunction are found in the literature. A first group of researchers, following Halliday and Hasan's (1976) early definition, have adopted a fairly narrow approach to conjunction, by making a very sharp distinction between logico-semantic relations holding between clauses, on the one hand, and relations holding between sentences or larger units, on the other hand. Those researchers consider the relations

of interpretative dependence between sentences (or larger units) to be the only significant instances of conjunction (see e.g. Halliday & Matthiessen 2004; Bloor & Bloor 2004; Matthiessen 2002; Eggins 1994). Accordingly, for them only an example such as (10), where the causal relation is expressed by means of an inter-sentential adverbial marker, is relevant for the study of cohesive conjunction. An example such as (11), by contrast, where the causal relation is signalled by means of a subordinator, is disregarded and considered as a linguistic phenomenon of a different nature (cf. below for more details). Likewise, while examples such as (12) and (14), where the concessive relation holds between two independent sentences, would be viewed as an instance of conjunction, Examples (13) and (15), where the very same relation, signalled by the very same items, holds between two clauses related by parataxis (see below for a definition), would not be viewed as cohesive.

(10)   It's raining. – **Then**, let's stay home.   (Halliday & Hasan 1976: 9)

(11)   **Since** it's raining, let's stay home.   (ibid.)

(12)   I personally favour the initiative and ardently support disarmament
         negotiations to reduce the risk of war. **But** I don't think endorsing a specific
         freeze proposal is appropriate for CCC.   (Halliday & Matthiessen 2004: 585)

(13)   We liked that breed of dog, **but** we felt we weren't in a position to own one
         at the time.   (ibid.: 407)

(14)   On dira que les Européens ne sont pas dépourvus de talents en matière
         d'espionnage industriel. **Mais** le fond du problème est qu'aucun d'eux
         ne possède individuellement une machine de la dimension d'Echelon.
                                                             (Rossette 2009: 14)

(15)   Leur monopole avait déjà été mis à mal par MBC, la célèbre chaîne
         saoudienne, pionnière en matière de télévision par satellite, **mais** la tornade
         qatarie a presque tout emporté.   (ibid.)

A second group of researchers – largely pioneered by Martin (1992) – have adopted a broader view of conjunction, questioning the clear dividing line drawn by Halli-day and Hasan between intra- and intersentential links (see among others: Martin 1983; Martin 1992; Gutwinski 1976; De Beaugrande & Dressler 1981; Martin & Rose 2007; Thompson 2014; Rossette 2009). In these accounts, the significance of sentence boundaries is disregarded on the grounds that "there are very close parallels to be drawn between intra-sentential and inter-sentential types of rela-tionship" (Butler 2003: 337). These researchers would include all examples in (10) to (15) in a study of conjunction.

    What fundamentally distinguishes the narrow from the broad definitions of conjunction thus has to do with the inclusion (or exclusion) of the system of taxis within the set of cohesive resources: while the advocates of a narrow definition draw a fairly sharp dividing line between the system of conjunction and the system

of taxis, those who adopt a broad definition consider hypotactic and paratactic links to be relevant instances of cohesive conjunction. Therefore, the selection of the broad or narrow approach in a study of cohesion is not simply a theoretical triviality but has a major impact on the design of the study itself. Depending on the approach adopted, the set of linguistic elements investigated will be markedly different: a study adopting a narrow approach to conjunction will overlook all subordinators as well as coordinators that are not used sentence-initially, as opposed to a study adopting a broad approach, which would include all these devices.[6] Figure 1 summarises the dispute animating the field of research on cohesive conjunction.

Narrow approach	Broad approach
Conjunctive cohesion Only includes inter-sentential logico-semantic relations	**Conjunctive cohesion** Both inter-and intra-sentential logico-semantic relations are part of the system of conjunction (i.e. taxis + conjunction)
Conjunctive markers Conjunctive adjuncts + sentence-initial coordinators	**Conjunctive markers** Conjunctive adjuncts + coordinators + subordinators (clause level or above)

Figure 1. Narrow and broad approaches to conjunction and conjunctive markers

Before I select one type of approach for the present study, it is necessary to describe parataxis and hypotaxis in greater detail, and to expand on the reasons that have led researchers to adopt each approach.

2.2.3.1   *Taxis in Systemic Functional Linguistics*
In Systemic Functional Linguistics, taxis refers to one of the two main systems through which clauses are related to one another to form clause complexes. It determines the type of dependency relationship that holds between two clauses and can be subdivided into two categories: (i) parataxis, which relates two clauses with equal syntactic status, as in (16) and (17); and (ii) hypotaxis, which binds

---

6.   Throughout this book, the terms 'coordinators' and 'subordinators' are used instead of 'conjunctions of coordination' and 'conjunctions of subordination'. The objective is to avoid any confusion that may arise from the homonymy between these two grammatical categories, and the key Hallidayan notion of cohesive conjunction, described above.

clauses of unequal syntactic status, with one clause being dependent on the other, as in (18) and (19).

(16)   She lived in a tiny bed-sit **and** she was not well-off **but** she entertained her guests in style.   (Butt et al. 2000: 165)

(17)   Moni était trouble et surpris **mais** il se laissait faire.   (Caffarel 2006: 30)

(18)   **Whereas** most children's fathers worked at an office, my father worked at the studio, so I went on the set.   (Halliday & Matthiessen 2004: 408)

(19)   Il était pâle comme Shiva, **alors que** alors que Vishnou, Rama, Krishna ont la peau sombre comme Suresh.   (Caffarel 2006: 32)

Importantly, while the distinction between parataxis and hypotaxis resembles the one made between coordination and subordination in traditional grammar, the categories do not fully overlap (Caffarel 2006: 23). Hypotaxis corresponds to the broad category of subordination excluding embedding (i.e. the process whereby a clause is 'rankshifted' to function as a group or part of a group in another clause).[7] Parataxis, on the other hand, is slightly more inclusive than the traditional category of coordination, as it also includes juxtaposition. In addition, in SFL some items which would not be analysed as coordinators, but rather as conjunctive adjuncts (or 'conjuncts', to use Quirk et al.'s 1985 terminology) in descriptive grammars are viewed as paratactic markers if they link two clauses in an equal relationship. This is the case in a sentence like (20), which is provided by Halliday and Matthiessen (2004) as an example of parataxis.

(20)   Through mounting irritation I kept telling him that I needed a cure for my son and nothing for myself; **still** I answered his questions with all the politeness I could muster.   (Halliday & Matthiessen 2004: 414)

The system that interacts with taxis to form clause complexes is a semantic one. It is referred to as the system of 'logico-semantic relations' and specifies the type of semantic relationships that hold between two clauses linked through taxis. Again, two main categories of logico-semantic relations may be identified. The first one is projection, and occurs when one clause serves to quote or report speech or ideas conveyed by another clause (Eggins 1994: 259). Projection is not relevant for the type of semantic relations investigated in the present book, however. The units analysed here (i.e. markers of contrast) rather pertain to the second category of logico-semantic relations, i.e. expansion, which occurs when "a clause is linked or bound to another clause in order to develop it, by explanation, addition, description, etc."

---

7.   Embedding will be defined in greater detail in Chapter 6.

(Bloor & Bloor 2004: 282). Expansion includes different kinds of circumstantial links (e.g. time, manner), but also such relations as causality, adversativity, addition or exemplification. The broad system of expansion is further subdivided into the three categories of extension, enhancement and elaboration, which, interestingly, are precisely the ones used by many Systemic Functional linguists to cut up the semantic space covered by cohesive conjunction (see below for a discussion of these similarities). The two types of taxis may combine with the different logico-semantic relations in order to form various kinds of clause complexes, as exemplified in Table 1 (although, notably, the various possible combinations are not equally likely, as clause complexing has been described as an "inherently probabilistic system"; see Nesbitt & Plum 1988; Matthiessen 2002; Caffarel 2006: 51–56).[8]

Table 1. Basic types of clause complex (Halliday & Matthiessen 2004: 380)

		Paratactic (1, 2, 3, etc.)	Hypotactic (α, β, γ, etc.)
(1) Expansion	**(a) Elaboration (=)**	1 John didn't wait; = 2 he ran away	α John ran away, =β which surprised everyone
	(b) Extension (+)	1 John ran away + 2 and Fred stayed behind	α John ran away, +β whereas Fred stayed behind
	(c) Enhancement (x)	1 John was scared x2 so he ran away	α John ran away, xβ because he was scared
(2) Projection	**(a) Locution (")**	1 John said: "2 I'm running away	α John said "β he was running away
	(b) Idea (')	1 John thought to himself : '2 I'll run away	α John thought 'β he would run away

It is also important to note that clause complexing is a recursive process, which means that more than one pair of clauses may be combined through parataxis and/or hypotaxis to form a clause complex. One example of a clause complex made up of more than two clauses is Example (16) above.

2.2.3.2   *Arguments in favour of the narrow and the broad approaches to conjunction*

This section lists the main reasons that have led researchers to adopt a broad or a narrow definition of conjunction. For the sake of clarity, when the term 'conjunction'

---

8.   It is beyond the scope of this study to expand on this classification of clause complexing; a detailed overview can be found in Halliday & Matthiessen (2004: chap. 7).

is meant in a narrow sense (i.e. as opposed to parataxis and hypotaxis), it is always preceded by 'intersentential'. The main argument advanced by the advocates of a narrow definition pertains to the distinction between grammatical and semantic linguistic resources. As explained in Section 2.2.1, cohesion in general is defined by Halliday and Hasan (1976) as a semantic instead of a structural type of relation, designed to account for linguistic links that may not be described by reference to grammar. This feature of cohesion is the one that has caused some researchers to draw such a clear-cut distinction between cohesive and tactic relationships, considering these two types of resource as being of a different nature. The difference is stated quite clearly by Eggins (1994: 279): "[i]n clause complex analysis, we analyse the structural relationships between clauses within sentences, while in conjunctive cohesion we analyse the non-structural relationships between sentences within text. [Therefore] the domain of clause complex relations is different from the domain of conjunctive cohesion." In other words, logical relations between sentences are purely semantic: they are not governed by any grammatical or structural imperatives (see e.g. Halliday & Hasan 1976: 6; Matthiessen 2002: 239, 263). They are viewed as the only significant instances of cohesion because they represent the variable aspect of cohesion (Halliday & Hasan 1976: 8; see also Halliday & Matthiessen 2004: 539). Hypotactic and paratactic relations holding between clauses, by contrast, operate primarily at the level of grammar, which implies that they are necessary for the grammatical correctness of the sentence in which they appear. They are not viewed as significant instances of cohesion because their use does not constitute a choice on the part of the writer/speaker.

A second argument in favour of the distinction between inter- and intrasentential relations pertains to the functions that the two types of resource perform in language (see e.g. Matthiessen 2002: 239; Crismore et al. 1993: 49; Eggins 1994): whereas intersentential links are said to belong to the textual metafunction of language (i.e. the dimension which has to do with the organisation of linguistic units to build a text), the devices ensuring the creation of clause complexes are said to contribute to the logical metafunction, which is itself part of the ideational metafunction (i.e. the function that has to do with the transposition of human experience in linguistic terms). Cohesion and taxis therefore operate at different levels of the linguistic system. These elements have led to the conclusion that "[cohesive] conjunction is not simply coordination extended so as to operate between sentences" (Halliday & Hasan 1976: 238).

By contrast, the main motivation that has led some researchers to broaden the category of cohesive conjunction by including tactic resources is the striking functional similarities between the two sets of linguistic devices. Butler (2003: 358) highlights the continuity between intra- and intersentential links by providing three examples extracted from the British National Corpus:

(21)   We had a great feast and lots to drink **and afterwards** there were recitations and a great singsong, and those that wanted to played poker.

(BNC HTG 621)

(22)   All the lovely prickles and twists stopped and she was left with nothing but a feeling of panic, able to think of nothing but how she could stop him without making him angry. **And afterwards** there was just a feeling of let-down, of wanting him to hold her and kiss her and pet her like a little girl.

(BNC BMW 2295–2296)

(23)   It was a noisy but not a violent affair, the clashes between the Blackshirts and the Red Front outside being mainly verbal. **Afterwards** there was a party at a Mayfair house, to which Joyce ensured that the visitor was invited.

(BNC EDA 1170–1171)

According to the narrow approach to cohesion initiated by Halliday and Hasan (1976), only Examples (22) and (23) should be viewed as cohesive, whereas the relation in (21) is considered as a phenomenon of a very different nature. As rightly observed by Butler, however, although there is undeniably a cline here in terms of the amount of work performed by grammar in the realisation of the logico-semantic relation, from a functional point of view it does not make much sense to make such a clear-cut distinction between (21), on the one hand, and (22) and (23) on the other. The effect achieved (i.e. relating two linguistic segments by temporality) is remarkably similar in all three examples. Christiansen (2011: 166) makes a similar point: he states that it can be misguided to lay too much stress on the distinction between intrasentential and intersentential semantic relations, one obvious reason for this being that in many cases, the only difference between an instance classified as cohesive, and another one considered as non-cohesive pertains to the use of punctuation (with one example containing a comma whilst the other includes a full stop, for instance). Granting so much discriminating power to punctuation is problematic and somewhat arbitrary, notably because the use of punctuation can be very erratic and idiosyncratic (Christiansen 2011: 27).

The functional closeness between intra- and intersentential linking devices is also very clear from the fact that, as explained by Martin (2001: 38) the semantics of intersentential conjunction may be aligned with the semantics of the logical metafunction.[9] As already mentioned earlier, the system of logico-semantic relations used for the description of intersentential conjunction is the same as the one developed to analyse paratactic and hypotactic relations of expansion: the same broad distinction is made between elaboration, extension and enhancement,

---

9. This has led some researchers to conclude that intersentential conjunction actually expresses logical meanings within the textual metafunction (Matthiessen 2002; Butler 2003).

with each of these categories being subdivided into a number of more specific semantic relations. Thus, each given logico-semantic relation may be realised either at the textual level, by intersentential conjunction, or at the logical level, by clause complexing (Matthiessen 2002: 282–283).[10] Table 2 illustrates the semantic correspondence between the systems of conjunction (defined in a narrow sense) and clause complexing, with examples of paratactic, hypotactic and intersentential conjunctive markers for each category within the logico-semantic system.

Table 2. Correspondence between the systems of logico-semantic relations of clause complexing and conjunction (adapted from Matthiessen 2002)

Type of relation	Clause complexing		Conjunction (intersentential)
	Hypotactic	Paratactic	Cohesive
Elaboration (apposition, clarification)	which, who [relative pronouns]	i.e, viz., e.g.	for example, for instance, in other words, I mean
Extension (addition, adversativity, variation)	while, whereas, besides, instead of	and, nor, but, or	also, neither, however, alternatively, on the other hand
Enhancement (matter, manner, spatio-temporal, causal-concessive)	before, after, if, because, unless, so that, (in order) to, in case of, despite, by	(and) then, so, for	afterwards, earlier, meanwhile, therefore, consequently, in this respect

The functional similarities between taxis and intersentential conjunction are in fact acknowledged even in the work of most proponents of the narrow approach to cohesion. In his article on clause complexing, for instance, Matthiessen (2002) constantly insists on the fact that conjunction (in a strict sense) and clause complexing are neighbouring systems in language, and that the boundaries between them are highly permeable. Halliday and Matthiessen (2004: 538–539) describe intersentential conjunction as a system that "has evolved as a complementary resource [to taxis] for creating and interpreting text [and] provides the resources for marking logico-semantic relationships that obtain between text spans of varying extent, ranging from clauses within clause complexes to long spans of a paragraph or

---

10.   This is in fact one of the main tenets of Rhetorical Structure Theory, a framework developed in close parallel with SFL, and which postulates that all rhetorical relations can be modelled on the basis of the relations holding in clause complexes (see e.g. Mann & Thompson 1988; Matthiessen & Thompson 1988).

more". As they explain, intersentential conjunction and taxis do the same thing, but at different linguistic levels: "[t]hese two resources complement one another in the grammatical realisation of transitions in text. The general principle of complementarity is this: clause complexing does relatively more work locally, while intersentential conjunction does relatively more work non-locally and even globally" (ibid.: 583). Finally, Halliday and Hasan (1976: 9) themselves, although they choose to disregard intrasentential relations of cohesion, concede that "cohesion is not, strictly speaking, a relation 'above the sentence'": since cohesive relations are fully independent of grammar, "[they] have in principle nothing to do with sentence boundaries" and "may be found just as well within a sentence as between sentences" (ibid.: 8). They explain that all in all, the main difference between inter- and intrasentential cohesion is simply that cohesive relations within the sentence "attract less notice because of the cohesive strength of the grammatical structure; since the sentence hangs together already, the cohesion is not needed in order to make it hang together" (ibid.).

The striking functional and semantic similarities between taxis and intersentential conjunction have led a number of researchers to conclude that the clause is a more relevant concept than the sentence for the study of cohesion. For the champions of a broad approach to conjunction, Halliday and Hasan's early account "fails to bring out the continuity between the structural […] and non-structural […] resources" available to link textual segments (Martin 1992: 19), and obscures the fact that "the alternative realizations are all variations on the same theme – namely that of relating one part of a text to another in terms of the natural logic of time, cause, comparison and addition" (ibid.: 163). Starting from this general observation, taxis has been integrated into conjunction in a multiplicity of ways: Martin (1992), for example, has designed a new, much broader framework of analysis of conjunction taking hypotactic clause complexes as its point of departure; Thompson (2014: 225) insists that taxis should be included in the category of conjunction, but not to that of cohesion; Caffarel (2006), on the other hand, does not mention cohesion or conjunction at all, but includes conjunctive markers (e.g. *cependant, toutefois, pourtant*) in her description of French clause complexing. It is beyond the scope of this study to provide a detailed overview of each specific description of conjunction.[11]

---

11.   Note that the debate surrounding the definition of cohesion and conjunction is visible through the whole body of research on conjunctive markers, whether or not the studies are explicitly grounded in the framework of cohesion studies and Systemic Functional Linguistics. While some researchers have grouped all the devices listed above into one broad category of functionally-equivalent devices, others tend to draw a line between the members of those various syntactic categories on the grounds that they display divergent linguistic features (see e.g. Touratier 2006 for a description of this debate in French linguistics).

2.2.3.3   *Other approaches to conjunction*

So far, Section 2.2.3 has focused on the two most commonly adopted approaches to the study of cohesive conjunction, viz. the approach that is restricted to the study of relations at intersentence level or beyond, referred to as the narrow approach; and the one that also includes paratactic and hypotactic relations holding between clauses, called the broad approach. It is worth mentioning, however, that the broad approach could be further subdivided into three 'sub-approaches' to conjunction. The one that has been discussed so far, and which looks at logico-semantic relations between clauses as well as sentences, is in fact the narrowest of three. Another approach to conjunction, adopted by de Beaugrande and Dressler (1981), for example, consists in also considering the relations occurring at the level of the phrase, such as the additive relations in (24) and (25), as relevant instances of conjunction (which de Beaugrande and Dressler call 'junction').

(24)   A great black and yellow rocket stood in a desert.

                                         (de Beaugrande & Dressler 1981: 51)

(25)   Hier matin, lui et moi avons discuté près du feu.

                     (Luscher quoted in Touratier 2006: 23; non-SFL framework)

Yet other researchers emphasise the importance of logico-semantic relations realised through logical metaphor for the study of conjunction. Logical metaphor occurs when "conjunctions are reconstrued as other kinds of elements, including processes, things, qualities and circumstances" (Martin & Rose 2007: 148). A typical instance of logical metaphor would consist in the reduction of a pair of clauses or sentences to a single clause, with the logico-semantic relation expressed as a process (with a verb phrase), and the two related clauses or sentences realised as nominal groups. One such example is provided by Caffarel (2006: 56; Example (26)), together with its clause complex equivalent (27). Here, a causal relation is expressed by the verb *entraîner*, and the processes in the two related clauses in (27) are reduced to two nominal groups built around *engouement* and *naissance*.

(26)   L'engouement des asparas, des danseuses du ciel, pour un homme peut entraîner la naissance d'un enfant. [The infatuation of asparas, dancers of the sky, for a man can result in the birth of a child].

(27)   Lorsque les asparas, des danseuses du ciel, s'éprennent d'un homme, il arrive qu'elles aient un enfant. [When the asparas, the dancers from the sky, fall in love with a man, it may happen that they have a child].

Other typical instances of logical metaphor would include verbs such as *to cause, to differ from, to result from*, etc. but also prepositions such as *because of, in spite of, due to* + noun phrase, or even nouns such as *conclusion, consequence,* etc. By

virtue of the fact that these phenomena take part in the construction of the logical structure of texts to a similar extent as intersentential conjunction, parataxis and hypotaxis, some researchers have suggested that they should also be included in studies investigating conjunction (see Martin 1992: 159–170 for a discussion; see also Moreno 1998, in a non-SFL framework).

These two alternative approaches are less widely adopted than the two approaches discussed earlier, which explains why they are only briefly introduced here. Figure 2 summarises the various approaches to conjunction found in the literature.

Narrow   **Broad**

| Intersentential relations only (e.g. Halliday & Hasan 1976; Eggins 1994; Halliday & Matthiessen 2004) | Intersentential relations + hypotaxis & parataxis (e.g. Martin 1983; Butler 2003) | Phrasal links (e.g. de Beaugrande & Dressler 1981); logical metaphors (see Martin 1992) |
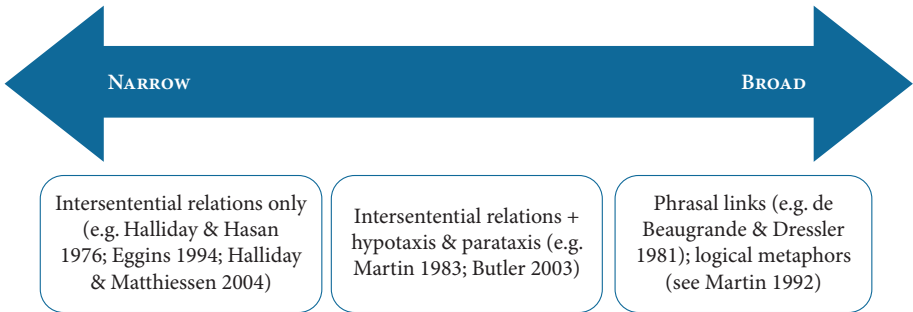
Figure 2. Approaches to cohesive conjunction

2.2.4   Towards a broad definition of conjunction

As explained in the previous section, the linguistic resources encompassed by conjunction in a narrow sense, on the one hand, and parataxis and hypotaxis, on the other hand, are functionally very close, in that they all serve to explicitly signal the logico-semantic relationships that hold between two discourse segments, within the same semantic space. As a result, theoretically, they can be used more or less interchangeably: a writer who would want to express a concession between (i) the fact that "estimates of the soot produced by the fires vary" and (ii) the fact that we can nevertheless probably equate it to "about 500,000 tonnes a month" could choose to do it by resorting to either: (i) an intersentential conjunctive adjunct (as in (28)); (ii) a paratactic link (as in (29)); or (iii) a hypotactic device (as in (30)), and achieve roughly the same effect.

(28)   Estimates of the soot produced by the fires vary. **However**, it is probably about 500,000 tonnes a month.   (Thompson 2014: 38)

(29)   Estimates of the soot produced by the fires vary, **but** it is probably about 500,000 tonnes a month.   (ibid.)

(30)   **Although** estimates of the soot produced by the fires vary, it is probably about 500,000 tonnes a month.   (ibid.)

One direct consequence of the functional similarity between the three resources of conjunction (in a narrow sense), parataxis and hypotaxis is that the preferences for one or the other type of option are likely to differ across languages (Matthiessen & Thompson 1988: 317). References to the fact that this is indeed the case abound in the literature. James (1980: 113), for example, states that "while every language has at its disposal a set of devices for maintaining textual cohesion, different languages have preferences for certain of these devices and neglect certain others". Likewise, Halverson (2004: 562) observes that "in spite of the fact that many languages seem to have similar sets of connective structures with ostensibly similar functional capacities, it seems to be the case that languages make very different use of the potential within the set of structures". In a translation-based contrastive analysis of text organisation in French and Japanese, Takagaki (2011) demonstrates that equivalents of conjunctive markers used intersententially are not necessarily intersentential themselves but may very well occur at clause level. Other studies providing evidence that the ways to signal logico-semantic relationships between the parts of a text vary across languages, even when they are closely related, include – to cite just a few: Lamiroy and Van Belle (1995); Fabricius-Hansen et al. (2005); Cosme (2006, 2008a); Castagnoli (2009); Cartoni et al. (2011); Becher (2011); Dixon and Aikhenvald (2011); Kunz & Lapshinova-Koltunski (2014, 2015); Kaplan (1966) and the field of contrastive rhetoric. French and English, in particular, have frequently been said to display marked differences in the ways in which they tend to package information, in spite of the fact that they have very similar devices at their disposal to do so (e.g. Vinay & Darbelnet 1995: 234; Hervey & Higgins 1992: 49; Poncharal 2005; Cosme 2006; see Chapter 3 for a detailed overview of the literature on English and French). From this it follows that in order to signal a given logical relationship explicitly, one language may tend to prefer intersentential conjunction, whereas in the same context, another language would show a preference for hypotaxis or parataxis. Knowing this, and keeping in mind the comparability requirements discussed in Section 2.1, it seems problematic to only investigate one type of device made available by languages to signal relationships at discourse level: in view of the fact that English and French have often been claimed to use partially different tools to organise information in discourse, not taking tactic resources into account in the present study would most probably be akin to biasing the analysis in favour of one or the other language. In other words, adopting a broad definition of conjunction is essential in order to guarantee an equal treatment of the two languages.

For this reason, this research adopts a broad definition of cohesion and conjunction, expanding the category of conjunction as traditionally defined by Halliday and Hasan (1976) to also include paratactic and hypotactic devices. In practice, this means that this study is interested not only in conjunctive adjuncts

of contrast (such as *however, by contrast, toutefois, cependant*) and sentence-initial coordinators, but also in coordinators used to relate clauses, and subordinators (e.g. *although, while, alors que*). These elements together constitute the category of conjunctive markers in this study. For the sake of clarity, these more traditional grammatical labels will be used in lieu of conjunction, hypotactic and paratactic markers, in view of the polysemy of the term 'conjunction' (both within SFL – cf. the broad and narrow approaches – and between SFL and more traditional grammatical accounts). Nevertheless, in the corpus analyses I will be careful to distinguish between coordinators and CAs used to link full sentences in a conjunctive relation, and those used to relate clauses in a paratactic relationship.

The choice to adopt a broad definition of conjunction is supported by the fact that most studies dealing with discourse relations in a contrastive perspective have adopted a similarly broad definition. Examples include Halverson (2004); Castagnoli (2009); or, in a Systemic Functional perspective, Kunz & Lapshinova-Koltunski (2014); Lapshinova-Koltunski & Kunz (2014). Another sign of the relevance of adopting a broad definition of conjunction and conjunctive markers is the fact that in most non-SFL frameworks of analysis of discourse relations, the broad approach is also preferred, with subordinators being considered as fully relevant signals of coherence relations (see e.g. Mann & Thompson 1988 for Rhetorical Structure Theory; Fraser 1999; Prasad et al. 2008 for the Penn Discourse Treebank; Sanders et al. 1992 on the Cognitive approach to Coherence Relations; see also Danlos et al. 2015a for the French Discourse Treebank). Finally, the broad approach to conjunction seems to be consistent with the kind of approach to cohesion that is dominant in the French literature. As explained in the introduction of this section, so far the Systemic Functional framework has mostly been developed in relation to English, and few studies have yet applied it to the description of French. As a result, most of the references quoted in this section, as well as the debate between the advocates of the broad and the narrow approaches, predominantly pertained to the analysis of the English language. In one of the very rare Systemic Functional accounts of French conjunction that I could find, however, Rossette (2009) adopts the same broad approach as the one taken here. One way in which she justifies her definition of conjunction is by reference to what is usually done in French linguistics (2009: 13):

> [W]hile in French the term "connective" is frequently used to group conjunctions of subordination (e.g. *lorsque, quand*), conjunctions of coordination (*et, mais, or, ou*) and connective adverbs (*donc, cependant*), whatever their position in the sentence, SFL does not provide such an overarching term, and rather makes a clear distinction between cohesive relations operating between sentences, and structural relations occurring within the sentence [my translation].

Therefore, a broad definition of conjunction seems to be in line with the approach that prevails in the French literature.

The *tertium comparationis* adopted for conjunctive markers is thus a functional one, along similar lines as Moreno (1998), Connor and Moreno (2005) or Crible (2018): the background of similarity justifying the comparison of conjunctive adjuncts, coordinators and subordinators across languages is their functional equivalence. The choice of a functional TC ties in well with the fact that the category of conjunctive markers is generally described in the literature as a functional one, which groups elements originating from a variety of grammatical classes (see for example Schourup 1999; Touratier 2006; Castagnoli 2009; Halverson 2004), and it takes into account the fact that different languages may express the same function through different grammatical means. The adoption of a functional TC for the study of discourse relations is also in line with Chesterman's (1998: 151ff) contrastive functional rhetoric, which starts from a text's macrostructure, then seeks to examine the ways in which it can be realised in different languages.

The approach to conjunction adopted in the present book is a broad one, but it is not the broadest: it includes neither phrasal links, nor instances of logico-semantic relations expressed through logical metaphor. The decision to exclude phrasal links is supported by an overview of the literature in the more general field of study on coherence relations, where it is generally accepted that relations between segments that are smaller than the clause do not qualify as discourse relations (see e.g. Fraser 1999: 940; Sanders & Spooren 2007: 924; Hoek et al. 2018 on this). Thus, in the present study, only markers linking clauses or larger units are considered for analysis.[12] Logico-semantic relations expressed through logical metaphor (e.g. verbs such as *to contrast, to result from*; nouns like *difference, conclusion*) are disregarded because they are viewed as a phenomenon of a radically different nature from conjunction, hypotaxis and parataxis. Their signals originate from open grammatical classes (unlike conjunctive adjuncts, coordinators and subordinators) and frequently function as core, fully obligatory constituents of the clause or sentence in which they appear (e.g. verb phrase, subject, complements, etc.). Although, as demonstrated by Moreno (1998), such expressions may be of interest in a contrastive perspective, they are considered to lie beyond the scope of the present study.

2.2.5   Some core features of conjunctive markers

In order to complete the description of conjunctive markers, this section outlines some of the main features of the category of conjunctive markers. More precisely,

---

12.   See Chapter 6 for a discussion on the definition of the clause.

it focuses on two elements that are of particular interest for the present study, i.e. (i) the optionality of conjunctive markers; and (ii) the specific features of each type of marker, i.e. conjunctive adjuncts, coordinators and subordinators. Although the three types of markers perform the same linking function, and can frequently be used interchangeably to signal a given discourse relation, they each display some specificities which make the choice of one or the other type of marker a "meaningful" one (Thompson 2014: 186).

For this section, it was necessary to complement the Systemic Functional framework by including studies investigating discourse relations from a variety of perspectives. In this respect, it is interesting to mention the considerable terminological explosion when it comes to the study of discourse relations: depending on the perspective adopted, the items called 'conjunctive markers' in this book are referred to in the literature by a large diversity of terms among which: connectives, connectors, discourse markers, pragmatic markers, conjuncts, conjunctive adjuncts, cue phrases, discourse operators, discourse relational devices, etc. in English; connecteurs (logiques), charnières, marqueurs de discours, conjonctifs, mots de liaison, organisateurs textuels, etc. in French, with the various terms usually covering only overlapping ranges of linguistic devices. This study is perhaps guilty of adding yet another term to that long list, in accordance with the approach that it adopts.

2.2.5.1   *Optionality*

The first feature of conjunctive markers (henceforth: CMs) that is worth mentioning is the fact that they are optional elements. The optionality of CMs is meant not in a grammatical sense since, as discussed in the previous sections, markers of taxis, and especially subordinators, cannot be deleted without affecting the grammatical correctness of the sentence. Rather, CMs are optional in the sense that in many cases, in addition to the choice between a conjunctive adjunct (henceforth: CA), a coordinator or a subordinator, writers can choose to simply resort to the juxtaposition of the related units, thus leaving the coherence relation between the two segments implicit. The logical relations expressed by conjunctive markers can generally be inferred on the basis of the linguistic context and/or the reader's world knowledge, and CMs have often been described as mere 'signals' of coherence relations that are present in the text even in the absence of a marker (see e.g. Taboada 2006; Das & Taboada 2013; Zufferey 2016). In Example (31), for instance, the reader can easily interpret the relation holding between the two clauses as a causal one, even though no explicit signal is used. Likewise, in (32) no explicit conjunctive marker is necessary to grasp the relation of opposition between the two segments.

(31)   Max fell. Jack pushed him.   (Cartoni et al. 2011: 79)

(32)   The earlier conflict was a one-ocean war for the navy and a one-theatre way for the army; the latter was a two-ocean war for the navy and one of five major theatres for the army.   (Martin 1983: 66)

Coherence relations have also been demonstrated to be signalled by alternative means to CMs, such as mood, modality, transitivity, punctuation, degree of finiteness of verbal elements, etc. (see e.g. Martin 1992: 165–166; Taboada 2006; Das & Taboada 2013). Thus, in Examples (33) and (34), the causal relationship is not strictly implicit, as the non-finite verbal form and the relative clause, respectively, provide clues to the reader as to the nature of the relation (Zufferey 2016: 265).

(33)   Feeling tired, Mary left the meeting early.

(34)   Mary, who felt tired, left the meeting early.

Taboada (2006) demonstrates that in a corpus of newspaper articles, coherence relations are signalled by means of CMs in only 43.5% of the cases, while the rest of the relations are either signalled by alternative resources (including logical metaphors), or simply left implicit.[13] Other studies have uncovered up to 70% of implicit discourse relations (see Taboada 2009). In view of this, several researchers have described CMs as reading instructions or "guides for interpretation" of the relations holding between textual segments, provided by the writer as some kind of "courtesy to the reader" to facilitate their interpretation of the texts. Evidence in favour of a facilitating role for CMs is provided in a number of studies that have demonstrated a positive influence of the presence of CMs on both text processing and text representation: for instance, the presence of explicit CMs in texts has been found to reduce reading times, and to induce better answers to comprehension and recall questions as compared to texts or fragments containing fewer or no explicit signals (see e.g. Degand et al. 1999; Degand & Sanders 2002; Sanders & Spooren 2007).

While it is broadly accepted that conjunctive markers are optional linguistic signals, there nevertheless seem to be two ways in which that optionality can be restricted, both highly relevant for the present study. Firstly, the degree to which discourse relations may be left implicit seems to differ across languages – and this is in fact one of the main rationales behind the analyses carried out in this book. A number of studies have suggested that, in the same way as different languages tend to prefer different types of markers to express similar relations (see above), languages differ in the extent to which they need to signal discourse relations explicitly by means of conjunctive markers. In other words, as Chesterman (1998: 185) puts

---

13.   Taboada focuses on six coherence relations, viz. concession, circumstance, result, background, elaboration and summary.

it, with respect to the expression of coherence relations "languages seem to differ in two main ways: the implicit/explicit parameter, and (if explicit) the actual form of expression". With respect to English and French, for example, the dominant claim has been that French tends to signal the logical relations between discourse segments by means of explicit markers more frequently than English, which would tend to leave these relations implicit (see e.g. Vinay & Darbelnet 1995: 234). Consequently, translation from English into French would require adding explicit relational markers that are absent from the original; conversely, translating from French into English would lead to the deletion of a number of superfluous explicit markers. The difference between English and French will be explored in greater detail in Chapter 3. Such cross-linguistic contrasts in terms of cohesive explicitness have been formulated for many other language pairs, including for example: English and German (see e.g. Becher 2011; Kunz & Lapshinova-Koltunski 2014; Kunz & Lapshinova-Koltunski 2015), English and Swedish (see e.g. Altenberg 1998, 2007) and French and Japanese (e.g. Takagaki 2011).

Secondly, a number of studies have demonstrated that some discourse relations are more likely to be left implicit than others, which usually require an explicit signal to be interpreted successfully. Researchers have looked into the factors making a given relation more or less likely to be conveyed implicitly and have postulated a number of ways to account for the phenomenon. For example, Sanders (2005) has put forward the "causality-by-default" hypothesis, which posits that by default, readers will tend to assume a causal relation between two consecutive segments, making causal relations particularly likely to be left implicit. Murray (1997: 228), on the other hand, suggests that it is the difference between continuous and discontinuous relations that explains why some relations require explicit signals more than others:

> "[R]eaders have a bias toward interpreting sentences in a narrative as following one another in a continuous manner. As readers progress through a narrative, they assume that the events will follow in a linear fashion. And when this occurs, reading is relatively easy. Continuity can be conveyed easily via additive or causal relations. When a reader encodes a text event that is discontinuous in the absence of a marker or indication of the discontinuity, reading is more difficult. Examples of discontinuity are numerous and include reversions to an earlier setting or scene (such as a flashback), an abrupt topic change, a surprising turn of events, a character moving away from what he/she is currently doing, or a violation of an expectation created in the previous text.

This means that relations such as cause or addition are more likely to be left implicit than concession or conditionality, for example. More recently, some researchers have suggested that the distinctions made by Sanders (2005) and Murray (1997) are not sufficient to account for the implicit-explicit dichotomy and that

instead, the degree of cognitive complexity of the relation, which is itself related to its degree of expectedness, determines whether or not it needs to be signalled explicitly (e.g. Hoek & Zufferey 2015; Hoek et al. 2017). It is beyond the scope of this study to expand on these issues, but the fact that different kinds of relations require different degrees of explicitness should be borne in mind throughout the analysis. This is especially true as relations of contrast have sometimes been said to be difficult to process in the absence of an explicit signal (see e.g. Sanders & Spooren 2007), and within this category the different subtypes of contrast have been shown not to behave in the same way (Asr & Demberg 2012; see Section 2.3 for more details). A final point concerning the optionality of conjunctive markers is that sometimes, language and semantics interact to determine whether or not a given relation should be signalled explicitly: some studies have shown that in some cases, the types of relations that require explicit signalling are not the same across languages. For example, Takagaki (2011) shows that, while Japanese tends to leave relations of causality and temporality implicit more frequently than French, the opposite is true of adversative and additive relations. Thus, in addition to general principles governing the implicit or explicit signalling of discourse relations, language-specific factors once again seem to play a part.

2.2.5.2 *Specific features of the three types of conjunctive markers*

In Sections 2.2.3 and 2.2.4, we established that the three types of markers included in the present study, viz. conjunctive adjuncts, coordinators and subordinators, are equivalent to some extent in that they all perform the same function: they express an explicit logico-semantic relationship between two discourse segments. However, this equivalence is not absolute, as each type of device has its own specific ways of signalling such relationships. One difference is that, while CAs and sentence-initial coordinators express purely semantic links, relations signalled with subordinators and inter-clausal coordinators are not only semantic but also grammatical. This has been discussed already and I will not dwell on this again here. Instead, I will focus on other differences, which mostly pertain to the specific semantic and/or discourse features associated with each type of marker. More precisely, the three devices may be contrasted along the following dimensions: (i) scope of the marker; (ii) status of the segments; (iii) positional flexibility; (iv) degree of explicitness of the relation; and (v) degree of integration. Note that for each of these dimensions, the three types of devices are grouped differently: sometimes, coordinators and subordinators resemble each other and are different from CAs; sometimes it is subordinators that differentiate themselves from the other two types of markers.

The first dimension along which the three types of markers differ pertains to their scope and the types of units that they relate. In the previous sections, we discussed the fact that, while coordinators (at least the ones used intrasententially)

and subordinators typically link clauses, thus acting rather locally, conjunctive adjuncts generally relate larger units, i.e. sentences but also paragraphs, at a more global level. This has two implications. First, CAs generally link fully complete units – which may themselves contain coordinate or subordinate links; whereas coordinators and subordinators may link all sorts of clauses, including reduced ones, which involve the ellipsis of some elements such as the verb or the subject (Chalker 1996: 2–3). In addition, while with coordinators (used intrasententially) and, especially subordinators, the exact delimitation of the segments related is provided clearly to the reader, with more global links such as the ones signalled by CAs, it falls on the reader to identify "how far back in the text the relation ranges" (Martin 1983: 39; see also Matthiessen 2002: 288), which is sometimes unclear (e.g. just the previous sentence, a whole paragraph, or even a non-adjacent sentence further back in text?).

Another implication of the choice of one or the other linking device pertains to the status of the related segments. Unlike the other two types of markers, the use of a subordinator generally implies that the information in one of the two segments (i.e. the subordinate clause) is relegated to the background, and presented as less important than the information included in the other segment (Quirk et al. 1985: 919; Leech & Svartvik 2002: 142). The syntactic asymmetry discussed earlier is therefore matched by a semantic one. In addition, subordinators are generally used when the writer intends to present some information as wholly or partly known by the reader (i.e. as given): subordinate clauses typically encode no new information (Quirk et al. 1985: 919; Leech & Svartvik 2002: 142; Martin 1983: 39). Thus, considerations of information structure may play a part in the choice of the type of marker used to relate two discourse segments.

A third element that may influence the choice of one specific type of marker pertains to the positional flexibility, both of the markers themselves, and the segments that they relate. Firstly, while subordinators and coordinators are both fixed at clause boundary, most conjunctive adjuncts may move around in the sentence and be used not only sentence-initially, but also medially and finally (Quirk et al. 1985: 921). The mobility of CAs is exemplified in (35).

(35) a. He tried hard. **However**, he failed [initial]
       b. He tried hard. This time, **however**, he failed [medial]
       c. He tried hard. He failed, **however** [final]

Given that the placement of CAs has been demonstrated to emerge as a tool to create various rhetorical effects in terms of information and thematic structure (such as emphasising the theme or the rheme of the sentence; indicating the boundary between given and new information; see e.g. Altenberg 2006; Lenker 2011; Dupont 2015), this feature of CAs may also influence the choice of one type of

marker over the other. The issue of conjunctive adjunct placement will be tackled in Chapter 8. Also related to issues of thematic structure, the use of a subordinator makes it possible for the writer to choose in which order s/he is going to organise the related segments, whereas with coordinators and CAs the order of the segments is fixed (Martin 1983: 39; Quirk et al. 1985: 921; Chalker 1996: 3). Clauses introduced by a subordinator may be used either sentence-initially or sentence-finally (and sometimes even sentence-medially), as exemplified in (36). In that respect, subordinators allow for more thematic flexibility, and leave the writer free to decide what part of the message s/he wishes to emphasise (e.g. the end position generally receives more focus than the initial position).

(36)   a.   **Although** <u>you make many good points</u>, the line of argument is not always clear.

b.   The line of argument is not always clear, **although** <u>you make many good points</u>.   (Thompson 2014: 192)

The three types of markers have also been opposed in terms of the degree of explicitness with which they signal logical relations. In that respect, for instance, coordinators have often been said to express the relationship between two segments in a vaguer, more implicit way than subordinators (see e.g. Quirk et al. 1985: 1040; Leech & Svartvik 2002: 142). Martin (1983: 83), too, claims that since the networks of devices developed by languages to express logico-semantic relations are generally more fine-grained for subordinators than for coordinators and conjunctive adjuncts, subordinators generally allow for a relation to be expressed in more specifically explicit ways. These distinctions in terms of explicitness have later been demonstrated to be too coarse, however, and the degrees of explicitness associated with different types of relational markers have been shown to be best represented as various levels on a scale (Cosme 2008a: 108–111; Castagnoli 2009: 56), with not only the type of marker but also the features of the related segments (e.g. the degree of finiteness of the verb) playing a part in the degree of explicitness of a given signal. To that end, Castagnoli (2009: 56) has put forward a cline – represented in Figure 3 – to determine the degree of explicitness associated with each type of realisation. Examples (37) to (45) illustrate each category on the cline.
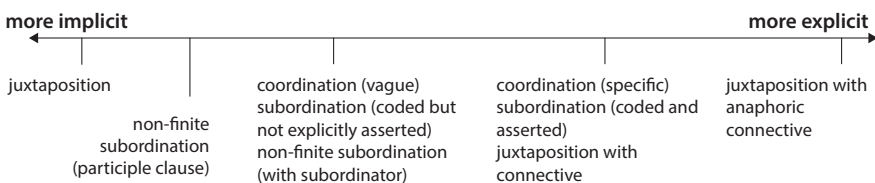
Figure 3. Cline of explicitness (Castagnoli 2009: 56)

(37)   *Juxtaposition*: The fields are green. It rained heavily.   (Castagnoli 2009: 54)

(38)   *Non-finite subordination (participle clause)*: <u>Being</u> a very gifted actress, she got the main part in the school play.

(39)   *Coordination (vague):*[14] I sent you three letters **and** got no answer.
(Castagnoli 2009: 56)

(40)   *Subordination (coded but not explicitly asserted):*[15] **After** the snow melted the river started swelling.   (adapted from Castagnoli 2009: 55)

(41)   *Non-finite subordination (with subordinator)*: **Though** <u>being</u> a terrible actress, she got the part in the school play.

(42)   *Coordination (specific)*: I sent you three letters **but** got no answer.

(43)   *Subordination (coded and asserted)*: The fields are green **because** it rained heavily.   (Castagnoli 2009: 54)

(44)   *Juxtaposition with connective*: I sent you three letters. **However**, I got no answer.

(45)   *Juxtaposition with anaphoric connective*: It rained heavily. **As a result of this**, the fields are green.

The present study only looks at openly contrastive markers, and therefore does not consider relations of contrast signalled by markers such as *and* or *when*, for example. As is clear from this figure, therefore, all the signals included here are equally explicit – apart from subordinators introducing a non-finite clause which, according to Figure 3, are less explicit than the other devices.

A final dimension along which coordinators, subordinators and CAs have been compared is the degree of integration of the relation expressed by the marker: as explained by Matthiessen (2002: 271), the different devices at our disposal to organise our discourse range from more to less integrated. In this respect, CAs have been said to express discourse relations in a looser way than coordinators and subordinators, which rather tend to signal the relation in a more integrated or compact fashion (Altenberg 1984: 38; Chalker 1996: 2–3). In an SFL perspective, Matthiessen (2002: 271–277) organises the different options on a cline ranging from more integrated – where one of the elements is presented as central and is

---

14.   In Castagnoli's work, vague markers seem to correspond to markers involving a semantic underspecification of the relation (e.g. in this case: *and* used to express a concession).

15.   In the same way as *and* was used to express a concessive relation in (39), here *after* is used instead of *because* to express a causal relation, leaving the reader to infer part of the relation uniting the two segments.

specified in some way by another element – to less integrated – where the two events or processes are represented as fully distinct entities, with each of them having their own thematic structure. In brief, Matthiessen presents hypotaxis as more integrated than parataxis, which is in turn more integrated than conjunction, while the juxtaposition of two sentences is the loosest type of connection. Each region on the cline is further subdivided into several levels, where finite and non-finite hypotaxis, for example, also differ in degree of integration (see also Lehmann 1988 for a similar cline of integration).

All these features may influence the choice of one type of marker over the others to signal a given logico-semantic relation in text. Despite the insistence, in this section, on the fact that the three types of linking devices have their own set of specific features, it should nevertheless be highlighted that the boundaries between the three categories are not always clear-cut (as is clear from the two clines of explicitness and integration just discussed): instead of three well-defined categories, what we have is different regions on a continuum, separated by fuzzy borders between which a number of borderline cases can be identified (see e.g. Quirk et al. 1985: 927; Lehmann 1988; Cosme 2008a: 108–111; Thompson 2014: 192; see also Chapter 7 for more on this).

2.3   Contrast

The identification and description of the relations that may hold between two segments of discourse is a hotly debated issue in discourse analysis, and the number, types and classifications of discourse relations vary, sometimes tremendously, across studies and frameworks. With respect to the number of relations identified, for example, a brief comparison of the main frameworks of analysis of discourse relations is particularly revealing: while Sanders et al.'s (1992) Cognitive approach to Coherence Relations (CCR) distinguishes between 12 main discourse relations, Segmented Discourse Representation Theory (SDRT; Asher & Lascarides 2003) identifies 15 relations, and the framework developed for the Penn Discourse Tree-Bank (PDTB; Prasad et al. 2008) puts forward a classification consisting of 30 different relations. In addition, although in its early form Rhetorical Structure Theory (RST; Mann & Thompson 1988) lists 23 main discourse relations, the reference manual later developed by Carlson and Marcu (2001) to annotate discourse relations within the RST framework lists no fewer than 78 distinct relations, spread across 16 categories.

This book focuses on one specific category of discourse relations, viz. the relation of contrast. In view of the high degree of indeterminacy affecting semantic descriptions at a very broad level, it is perhaps not surprising that, at a more

granular level of description, an overview of the literature on contrast also reveals the existence of highly varied, frequently conflicting ways of defining and categorising discourse relations of contrast – whether the studies are concerned with French, English, or pursue more universal descriptive goals. The lack of consensus characterising the definition of contrast is made clear by Fraser and Malamud-Makowski (1996: 865):

> The notion of contrast in English, and presumably other languages, is not well-defined. Definitions of contrast in the literature range across the semantic, logical, pragmatic, functional, and discourse domains, and some of these definitions overlap and intersect. Given this uncertainty over what seems to be a basic notion, it is hardly surprising that there is no agreement on what constitutes the category of contrastive discourse markers, if in fact a class exists.

In view of these difficulties in defining contrast, Fraser and Malamud-Makowski (ibid.) conclude that "[r]ather than spend time, perhaps fruitlessly, in search of a suitable definition, we will simply analyze a group of discourse markers which intuitively fall within the rubric of 'contrastive'". Unlike Fraser and Malamud-Makowski, instead of relying on my intuition to select the linguistic items to be analysed, I will strive to establish a well-documented, thoroughly-circumscribed definition of contrast that will be valid for the description of both English and French. The present section is devoted to this endeavour. Providing a reliable definition of contrast is felt to be a prerequisite to the identification of the linguistic items that may in fact be considered as markers of contrast (see Chapter 5), which is in turn a crucial stage for the validity of the corpus-based cross-linguistic comparisons carried out in Chapters 7 and 8. The definition of contrast formulated here will be taken to constitute the semantic part of my *tertium comparationis*, complementing the functional TC established for conjunctive markers.

As explained earlier, the Systemic Functional framework is particularly well-adapted to the purposes of this study, notably because of the considerable attention that it pays to cohesion and the mechanisms through which clauses and sentences are linked with one another to form texts. When it comes to the description of the *meaning* of conjunctive markers, however, the apparatus provided by SFL is not always helpful, as the semantic descriptions that it provides are often insufficiently detailed to get a firm grasp on the fine-grained distinctions between the various relations that may unite discourse segments. In their *Introduction to Functional Grammar*, for instance, Halliday and Matthiessen (2004) only provide short definitions for each of the broad categories identified (e.g. addition, variation, manner, etc.) and merely assign labels, accompanied by one or two examples, for each subcategory, leaving it to the reader to infer the difference between

meanings at more delicate levels of categorisation. For this reason, in this section I will complement the Systemic Functional accounts on the semantics of CMs with non-SFL descriptions of contrast in English and French, but also more general, universally-oriented descriptions of this relation.

One aspect that seems widely accepted in the literature is that contrast constitutes a general semantic category subsuming other, more specific ways of expressing a contrast between discourse units. Beyond this general common ground, there is considerable variation across studies. More particularly, a careful overview of the literature on contrast highlights three major dimensions along which descriptions of contrast differ, viz. (i) the number and types of relations of contrast identified; (ii) the way these relations are classified; and (iii) the labels assigned to both the general category of contrast and the subtypes of contrastive relations. Another difficulty inherent in the study of discourse relations of contrast is that definitions of contrast often remain vague in the literature. On the one hand, some researchers simply fail to provide an operational definition of contrast (cf. e.g. Fraser & Malamud-Makowski 1996, cited above). On the other hand, the comparison of approaches is made very hard by the fact that many studies only provide obscure or imprecise definitions of the subcategories that they identify within contrast, and simply provide one or two examples to illustrate the distinctions made. As the nuances between the subtypes are sometimes very subtle, it can be hard to work out the exact delimitation of the categories.

The objective of this section is to try to identify a definition of contrast for the present study based on the diversity of descriptions available. Section 2.3.1 first provides a brief state-of-the-art overview of the research on contrast, trying to bring order to the multitude of approaches available. Against this backdrop, Section 2.3.2 puts forward a definition of contrast for the present study, and describes the different relations making up the category of contrast in greater detail. The definition of contrast adopted here was felt to be a faithful reflection of the descriptions found in the literature, while also reconciling accounts of contrast in French and English in the best way possible. Importantly, it should be borne in mind throughout this section that I have no ambition to provide a fine-grained classification accounting for all the semantic subtleties encompassed by the meaning of contrast. Rather, the objective is to clearly circumscribe the broad semantic area covered by this study. On this basis, I will be able to determine with as much certainty as possible which English and French markers (or senses of markers, for polysemous items) match (or do not match) that definition, and therefore should (or must not) be included in the corpus analysis. In addition, this section will be restricted to relations of contrast that are relevant for the study of written language. Contrastive relations that are specific to speech/dialogical language will not be discussed here.

2.3.1   Overview of the literature on contrast

This section provides a general overview of the definitions and classifications of discourse relations of contrast in the literature. It first presents the debate concerning the number of subtypes which should be identified within the general category of contrast. It then gives a snapshot of the variety of ways in which these relations are classified.

2.3.1.1   *Number and types of relations of contrast*

As stated above, in the literature there is a relatively broad consensus that contrast should be viewed as a general semantic category encompassing several, more specific subtypes. Beyond this basic premise, however, the approaches to contrast are extremely diverse. As pointed out by Mann and Thompson (1992: 42), "the most obvious difference between approaches is in how many relations are recognised". In this respect, apart from the SDRT framework, which identifies one single relation of contrast (Asher & Lascarides 2003; see also Muller et al. 2012 for the application of SDRT to French), most accounts make a minimal distinction between two categories within contrast. Nevertheless, already at this coarse level of subclassification, there is significant variation across studies. In fact, the basic dichotomy made within contrast constitutes the most striking difference between studies investigating contrast in English and French (see Foolen 1991: 82; Izutsu 2008: 647 on the differences between approaches to contrast across languages).

In English, the division that is most often made within contrast is the one between (i) semantic opposition and (ii) denial of expectation (Lakoff 1971; see also Blakemore 1989). The distinction originates from Lakoff's (1971) study of the coordinator *but* (which has often been described as "the quintessential contrastive […] marker"; Fraser & Malamud-Makowski 1996: 865) but was later broadly applied to the category of contrast in general. In relations of semantic opposition (hereafter opposition), the marker simply points to the existence of a difference between the two discourse segments that it relates.[16] Semantic opposition is exemplified in (46), where the marker *but* indicates a difference in height between John and Bill. Note that for greater convenience, from this point onwards the two segments related by CMs are referred to as S1 and S2, where S1 corresponds to the first segment linked by the CM, while S2 corresponds to the second segment in the relation, introduced by the CM.

---

16.   For the sake of clarity, I have taken the liberty to smooth over the terminological differences between studies in this section, grouping under a single label the relations that appeared identical, even if they were not referred to in the same ways. Issues of terminology are touched upon at the end of this section.

(46)   John is tall **but** Bill is short.   (Lakoff 1971: 133)

Denial of expectation – which has more frequently been renamed 'concession', a term which I will prefer here – occurs when the second segment in the relation (i.e. S2) rejects a pragmatic inference that the reader/listener may be taken to have drawn from the first segment (i.e. S1). In (47), for instance, S1 leads us to expect that Bill passed his exams, as it is often the case that when people work hard, they succeed. The second segment (S2), however, rejects that conclusion.

(47)   Bill studied hard, **but** he failed the exam.   (Izutsu 2008: 649)

Accounts of contrast in French revolve around the central dichotomy, not between opposition and concession, but between corrective and non-corrective types of contrast. Similarly to what is observed in the English literature on contrast, this distinction was initiated by Anscombre and Ducrot (1977) in their semantic description of *mais*, and later taken up as a basis for the general description of contrast in French. In their seminal article, Anscombre and Ducrot distinguish between two uses of *mais*, which correspond to the lexical distinction made between *sondern* and *aber* in German, and between *sino* and *pero* in Spanish.[17] The first use, referred to as the SN (*sino/sondern*) use, occurs when S1 takes the form of a negative sentence, while S2 provides a correction of the information rejected by S1. In (48), for example, S1 refutes the information that Peter is French, and S2 provides the correct information, i.e. that he is in fact German.

(48)   Pierre n'est pas français, **mais** au contraire il est allemand [Peter is
        not French, **but** on the contrary he is German].
                                        (Anscombre & Ducrot 1977: 34)

The second use identified by Anscombre and Ducrot is called PA (*pero/aber*), and seems to roughly correspond to the concessive use described in the English literature (see Riegel et al. 2001: 619): it is described by Anscombre and Ducrot (1977: 28) as signalling a relation in which S1 points towards a given conclusion, which S2 then rejects. One example of the PA use of *mais* is provided in (49): as was the case in (47), here the conjunctive marker indicates that S2 rejects a conclusion that may be drawn from S1, viz. that since he is a Republican, he must be dishonest.

(49)   Il est républicain **mais** il est honnête [He is a Republican, **but** he is honest].
                                        (adapted from Anscombre & Ducrot 1977: 34)

---

17.   This lexical distinction is also found in other languages, although they are not mentioned in Anscombre & Ducrot (1977). See Foolen (1991) for more details on this.

For greater transparency, the SN use will henceforth be referred to as correction, whereas the PA use will be called concession. The corrective/non-corrective dichotomy introduced by Anscombre and Ducrot lies at the basis of most studies on contrast in French. Riegel et al. (2001: 620), for instance, also endorse it as the basic distinction with respect to contrast, stating that "other terms expressing [contrast] correspond, with distinctive nuances, to these two values of *mais*" [my translation]. Thus, in French, other types of contrast – including opposition – are frequently treated as peripheral (Stoye 2014: 49).

In summary, the basic distinction made within contrast is not the same in research on English and French. As a matter of fact, the lack of equivalence between the French and English accounts has proved puzzling for some researchers, who have confused the two distinctions (Foolen 1991: 84). Blakemore (1989: 15), for example, writes the following:

> Many writers have found it necessary to distinguish between two uses of *but*: the so-called 'denial of expectation' use […] and the so-called 'contrast' [i.e. semantic opposition] use […]. R. Lakoff (1971) presented this distinction as a distinction between two meanings of *but*, a proposal which would seem to find support in the fact that in some languages (for example, German, Spanish and Hebrew) *but* may be translated by either of two words. Indeed, as Anscombre and Ducrot (1977) and Horn (1985) have argued, although the distinction is not realized lexically in languages like French and English, it gives rise to the same differences in distribution that distinguish *aber* and *sondern* in German and *pero* and *sino* in Spanish.

Here, the distinction between 'denial of expectation' and 'semantic opposition' is mapped onto the one between the 'SN' and the 'PA' subtypes of contrast. The same type of mix-up is found in Salkie & Oates (1999: 35). As I hope to have shown clearly, however, the two distinctions are not equivalent: although the 'PA' use identified by Anscombre and Ducrot (1977) broadly corresponds to concession (or 'denial of expectation') as defined by Lakoff (1971), correction and opposition are different types of relation.

In addition to the studies viewing contrast as a binary relation, a number of studies have made further distinctions within the general category of contrast. Firstly, in the literature we find both accounts of contrast in English that have integrated the relation of correction (e.g. Halliday & Hasan 1976; Martin 1983; Martin 1992; Fraser 1998), and studies on French that have incorporated the relation of opposition (e.g. Csüry 2001; Charaudeau 1992; Van de Voorde 1992; Danlos & Roze 2011; Stoye 2014). This goes to show that the two sets of distinctions discussed earlier seem to be equally valid in English and French, despite the fact that each tradition has tended to lay more focus on one of them. Moreover, interestingly many studies aiming to provide a description of relations of contrast

across several languages, or which have the ambition to provide a universally-valid description of contrastive relations, also adopt the tripartite view of contrast distinguishing between opposition, concession and correction. Notable examples include Foolen (1991) on English, German, Russian and Hebrew;[18] Izutsu's (2008) language-neutral description of relations of contrast; Rudolph's (1996) study of contrastive relations in English, German, Spanish and Portuguese;[19] Lamiroy and Van Belle's (1995) study of contrast in French and Dutch, or Taboada and Gómez-González's (2012) study on English and Spanish. Likewise, Rhetorical Structure Theory (Mann & Thompson 1988), whose objective is to provide a language-neutral description of discourse relations, identifies three relations within contrast, which seem to broadly correspond to the relations of opposition, concession and correction described above (see also Mann & Thompson 1992 for an article dealing specifically with contrast).[20]

In addition, alongside these three seemingly central relations, a number of other meanings have been associated with contrast in the literature. An overview of current research brought out three main additional relations which are also associated with contrast, viz. exception, reformulation and dismissal. The relation of exception applies when the CM "indicates that [S2] specifies an exception to the generalization specified by [S1]" (Prasad et al. 2008: 37). In Example (50), S2 restricts the scope of S1 by specifying an exception to it.

(50)   The boy was alright **except that** he got pretty wet.   (Martin 1983: 34)

In relations of reformulation, the conjunctive marker indicates that the writer/speaker wishes to replace what s/he has just said by a different – clearer, more accurate or more important – formulation. In Example (51), Alice rephrases what she has just said to provide a revised, more accurate version of it.

(51)   'What a beautiful belt you've got on', Alice suddenly remarked… '**At least**', she corrected herself on second thoughts, 'a beautiful cravat, I should have said, not a belt'.   (Halliday & Hasan 1976: 254)

Finally, relations of dismissal signal that some element previously mentioned in discourse should in fact be dismissed as irrelevant (Halliday & Hasan 1976: 255).

---

18.   Foolen actually views these three meanings as embodying pragmatic differences, pertaining to polyfunctionality, rather than genuine semantic differences.

19.   Rudolph also includes additional distinctions. However, as they pertain to speech-specific uses of contrastive markers, they are not discussed here.

20.   This is what I was able to gather from their description of discourse relations. However, as they frequently provide very abstract definitions of their relations, generally accompanied by a single example, I hope that my interpretation of their three categories is accurate.

By means of the marker *in any case* in (52), for example, speaker B signals that his/her initial answer 'I'm not hungry' is after all irrelevant: even if s/he were hungry, s/he still would not eat dinner, as s/he has plans to go out to eat.

(52)   A: Dinner's ready.
         B: But I'm not hungry. **In any case**, I'm going out to eat.   (Martin 1983: 35)

A final semantic distinction that deserves mention here is the one between internal and external discourse relations. Some researchers stand out by making a crucial distinction between internal and external ways of expressing contrast, which emerges as a criterion for identifying further subtypes of contrast (see e.g. Halliday & Hasan 1976; Martin 1983; Martin 1992; Oversteegen 1997; Prasad et al. 2008). This distinction is in fact a central one in the field of study of discourse relations in general, and the external vs internal opposition has also been referred to as 'semantic vs pragmatic' (Van Dijk 1979; Oversteegen 1997); 'subject matter vs presentational' (Mann & Thompson 1988); 'objective vs subjective' (Pander Maat & Degand 2001); 'content vs epistemic and speech act' (Sweetser 1990), etc. The basic distinction is the following: external relations hold between events or entities in the outside world, and therefore structure information at the experiential level of language; internal relations have to do with the organisation of textual units instead of real-world events, and rather pertain to the interpersonal linguistic metafunction. They serve to structure the arguments of a text in a logical way (see Halliday & Hasan 1976: 238–44; Martin 1992: 178–84 for more details). The distinction between external and internal relations is very clear for temporal relations, for instance. In Example (53), the marker establishes a temporal link between two events that occurred in the outside world, in order to specify the order in which they took place. In Example (54), by contrast, the temporal markers are used to organise textual content in a way that will make the steps of the writer's argumentation clear to the reader. The segments related do not refer to events that follow one another in the real world. An example of the internal/external distinction for contrast is provided in Examples (55) and (56) below, where (55) illustrates an external contrast, whereas (56) illustrates an internal one.

(53)   I was told to shut up, sit in a chair, **then** I was questioned.
                                                                (Martin & Rose 2007: 126)

(54)   **Firstly** the act required that where the offence is a gross violation the application should be dealt with in a public hearing […]. **Secondly** it is not true that amnesty encourages impunity because amnesty is only given to those who plead guilty […]. **Finally**, retributive justice […] is not the only form of justice […] there is another kind of justice, restorative justice.
                                                                              (ibid.: 138)

(55)   The deer threw the boy in the pond **whereas** the dog fell on its own.

(Martin 1983: 23)

(56)   On the one hand we could view such grammars as false. **On the other hand** it might be preferable to see them as incomplete.   (ibid.: 31)

In summary, this section has listed and defined the main subtypes of meanings that have been associated with contrast in the literature. Six main relations (i.e. opposition, concession, correction, exception, reformulation and dismissal) were identified, and reference was also made to studies adding a distinction between the internal and external versions of some of them. As explained earlier, all these sub-types were not mentioned in all studies, and significant variation was found in the number and types of relations identified across studies. In this respect, three types of situations are observed: (i) in some studies, (some of) the relations described above are clearly described as relations of contrast; (ii) in other studies, these rela-tions are identified as relevant discourse relations, but are associated with another broad category, such as addition or consequence. These issues of classification are discussed in Section 2.3.1.2. Finally, (iii) in some studies, some of the relations discussed in this section are simply not mentioned at all. Table 3 provides a sum-mary of some of the main approaches to contrast found in the literature in terms of the number and types of relations identified, in order to clearly show the lack of consensus in that respect. The cells of the table are left blank when the relation is not identified at all by the researcher(s); they are ticked when a given relation is both identified and associated with contrast; and a circle indicates that the relation is identified but not directly associated with contrast. An explanatory footnote is added whenever additional information was felt to be necessary.

It should once again be noted that in accordance with the goals of the pres-ent study, this overview of contrastive relations has remained fairly general. Many additional, fine-grained distinctions within the categories discussed here are also found in the literature. Concession, in particular, has given rise to a multiplicity of studies trying to disentangle the subtle differences in meaning between differ-ent (uses) of concessive markers (see e.g. Morel 1996; Grote et al. 1997; Altenberg 2002; Rudolph 1996; Mortier & Degand 2009). It is beyond the scope of this study to go into any detail about these distinctions.

As appears from this table, even though no consensual description of rela-tions of contrast can be found at first sight, when one goes through the painstak-ing task of mapping the different descriptions onto one another, it is possible to identify at least some common ground between them. If that were not the case, I would not have been able to group several references in the same cell as is the case here, where the studies identifying both the same number and the same types of relations of contrast appear together. Even when such similarities exist however,

Table 3. Number and types of relations of contrast identified across studies

Language	Study/ies	Opposition	Concession	Correction	Exception	Reformulation	Dismissal
English	Lakoff 1971; Blakemore 1989; Biber et al. 1999	V	V				
	Fraser 1998*	V	V	V			
	Quirk et al. 1985**	V	V			V	
	Halliday & Hasan 1976; Christiansen 2011	O	V	V	V	V†	V
	Celce-Murcia & Larsen-Freeman 1999 Prasad	V	V	V		V	V
	et al. 2008	V	V	O	O	O	
	Martin 1983, 1992‡; Martin & Rose 2007	V	O	V	V	O	O
	Halliday & Matthiessen 2004□	O	O	O	O	O	O
French	Anscombre & Ducrot 1977∞; Riegel et al. 2001		V	V			
	Van de Voorde 1992; Csüry 2001	V	V	V			
	Danlos & Roze 2011	V	V	O■	O	O	O
	Charaudeau 1992; Caffarel 2006; Stoye 2014	V	V				

(*Continued*)

Table 3. (*Continued*)

Language	Study/ies	Opposition	Concession	Correction	Exception	Reformulation	Dismissal
Multilingual/ language-neutral accounts	Foolen 1991; Izutsu 2008; Taboada & Gómez-González 2012; Mann & Thompson 1988, 1992; Lamiroy & Van Belle 1995; Rudolph 1996	V	V	V			
	Salkie & Oates 1999	V	V				
	Sanders et al. 1992	V	V		V		

* Fraser (1998) groups the meanings covered by contrast and concession in a single category (which groups markers that "signal that the speaker intends the explicit message conveyed by S2 to contrast with an explicit or indirect message conveyed by S1" (ibid.: 306), and splits correction into two distinct categories, according to how the markers expressing each type of correction behave in dialogues.

** In Quirk et al. (1985), two different categories of reformulation are distinguished. In addition, their list of so-called 'antithetic conjuncts' (broadly corresponding to opposition markers) contains some markers of correction, but no explicit mention of this relation is made in the (otherwise vague) definition of antithesis.

† Some types of reformulation are associated with the additive relations.

‡ In Martin (1983, 1992), the category of contrast is included into the more general class of comparative relations, together with relations of similarity.

□ In Halliday and Matthiessen (2004), there is no general category of contrast per se. Instead, all the relations of contrast are scattered across the functional categories of elaboration, extension and enhancement (see 2.3.1.2 for more details).

∞Some researchers say that Anscombre and Ducrot's PA use of *mais* also includes the opposition subtype of contrast. However, although oppositive uses of *mais* are mentioned very briefly in Anscombre and Ducrot's article, they are clearly not central in the definition of *mais* PA, which rather revolves around the idea of a denied expectation.

Correction is included in the broad category of "non-veridical relations". Also note that in Danlos and Roze (2011), contrast is a subcategory within the broader class of relations of comparison. Finally, reformulation and dismissal are not mentioned explicitly, but are both included in the more general category of 'evaluation'.

they are frequently obscured by the considerable diversity of labels that are used to refer to roughly equivalent relations. The literature provides us with a large variety of terms to refer to both the general category of contrast, and its various subtypes. On the one hand, it is frequent to encounter different labels to refer to the same (set of) relation(s), sometimes even in the same book: in Riegel et al. (2001), for example, the term 'adversatif' is used to refer to coordinators expressing contrast, whereas 'opposition' and 'concession' are used to talk about conjunctive adjuncts. On the other hand, the same labels frequently refer to different concepts across studies. For example, the term 'contrast' is used to refer to (i) the general category of contrastive relations, as is the case in the present study (e.g. Quirk et al. 1985; Fraser 1998; Mann & Thompson 1992); (ii) the relation referred to as 'opposition' here (e.g. Izutsu 2008; Biber et al. 1999; Csüry 2001); and even (iii) a subset of concessive relations (Halliday & Hasan 1976; Christiansen 2011). This terminological confusion makes it all the more difficult to compare descriptions of contrast.

2.3.1.2   *Categorisation of relations of contrast*

In addition to divergences in the number and types of contrast identified, and in the labels used to refer to them, an overview of the literature reveals that the ways in which relations of contrast are classified also varies markedly across studies. In view of this diversity, here I will simply provide a few examples illustrating the types of divergences found in the literature, instead of trying to provide a comprehensive account of all the existing classifications of contrast. Fundamentally, two main tendencies emerge: while some researchers tend to group the relations identified in Section 2.3.1.1 in one single, superordinate category (e.g. Quirk et al. 1985; Fraser 1998; Van de Voorde 1992; Csüry 2001; Rudolph 1996; Izutsu 2008), in other studies the relations that are generally associated with contrast are scattered across several categories (e.g. Halliday & Hasan 1976; Halliday & Matthiessen 2004; Martin 1983; Martin 1992; Adam 2008; Danlos & Roze 2011). This appears very clearly when comparing, for example, Celce-Murcia and Larsen-Freeman's (1999) classification represented in Figure 4, with Halliday and Matthiessen's (2004) system of conjunction (Figure 5), in which all the relations described above are circled (although the labels are different from the ones I have used so far, see above on issues of terminology).[21] Whereas in Celce-Murcia and Larsen-Freeman's account,

---

21.   The correspondences with the labels used in this chapter are as follows: (i) corrective = reformulation; (ii) dismissive = dismissal; (iii) adversative = opposition and some types of concession; (iv) replacive = correction; (v) subtractive = exception; (vi) concessive = concession.

five of the six relations mentioned earlier are grouped in one single category (called adversative), in Halliday and Matthiessen those relations are disseminated across the three main categories (elaboration, extension and enhancement), and do not even appear at the same level of the taxonomy (e.g. adversative relations are provided at the second level of the classification [extending > adversative], whereas concession appears at the fourth level [enhancing > causal-conditional > conditional > concessive], despite its importance as a rather general category in many other frameworks).
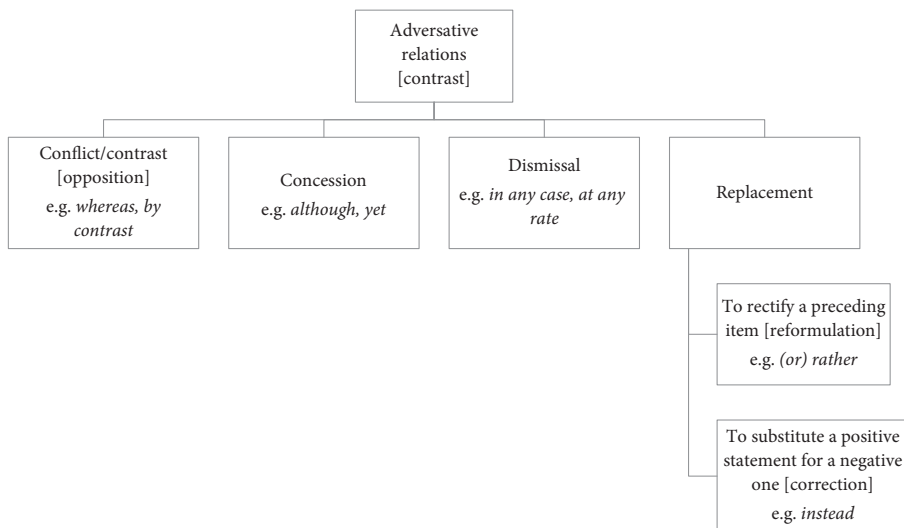
Figure 4. Celce-Murcia and Larsen-Freeman's (1999: 326) classification of relations of contrast

Another scattered view of relations of contrast is provided by Martin (1983, 1992). Martin considers most of the relations commonly associated with contrast (i.e. opposition, correction, exception and dismissal) in the broader context of comparison, and presents them as the counterpart of relations of similarity (expressed by such markers as *similarly, likewise, for example*), which, interestingly, include reformulation. Concession, on the other hand, is presented separately from the other five subtypes and associated with consequential relations. The idea is that concession can be viewed as a negative counterpart of causal-consequential relations, since what happens in concession is that the usual consequence of S1 is invalidated by S2 (see König & Siemund 2000; Verhagen 2000 on the relationships between concession and causality). In Halliday and Hasan (1976), on the other hand, relations of concession, correction, dismissal, and some types of reformulation are grouped in a single category labelled 'adversative', whereas opposition falls
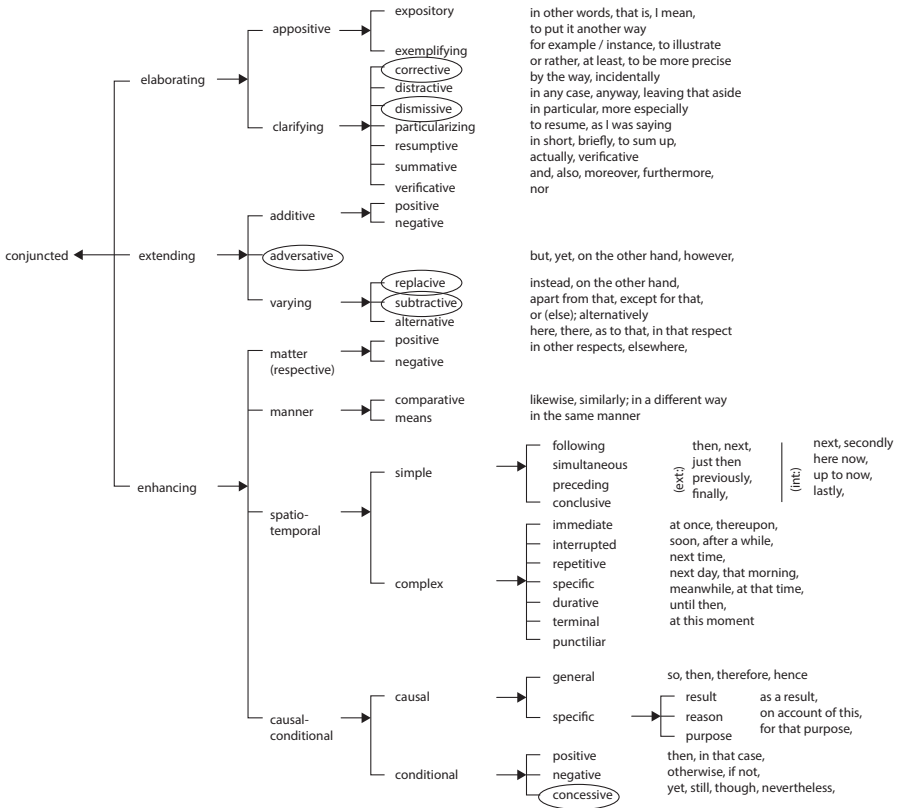
Figure 5. System of conjunction in Halliday and Matthiessen (2004: 541)

into the additive category, being viewed as a negative counterpart of relations of similarity (e.g. *likewise, similarly*, etc.).

Yet another, fairly widespread approach is adopted by Sanders et al. (1992), who attempt to define all discourse relations by reference to a set of four cognitive primitives whose levels can be combined to specify the various relations that may hold between two clauses or sentences. The four dimensions are the following: (i) basic operation, which postulates that all relations can be brought down to the addition vs causality opposition, depending on whether or not a relation of implication holds between S1 and S2; (ii) source of coherence, i.e. semantic vs pragmatic, which corresponds to the external vs internal distinction discussed earlier; (iii) order of the relation, i.e. basic vs non-basic, which basically refers to whether the segments are ordered in discourse in the same way as they would occur in the real world (for example, a basic order would correspond to a sequence where the cause precedes the consequence, whereas a relation with a non-basic order would present the consequence before the cause, as in: *She got wet because it rained*); (iv) polarity, i.e. positive vs negative, which depends on whether or not the

relation between the two segments involves the negation of some of their content, or what they imply; in this sense, for instance, concession is considered as a negative relation, since it leads to the negation of what follows from S1. In Sanders et al.'s framework, concession is separated from the other contrastive relations (e.g. opposition, exception) in terms of basic operation: it is viewed as a causal relation, whereas the other contrastive relations are classified as additive. On the other hand, all the contrastive relations identified in this framework are grouped with respect to the polarity dimension, as they are all considered to be negative relations.

It would be possible to write dozens of pages on the discrepancies between classifications of contrastive relations. However, I hope that this brief summary has demonstrated clearly the extent of the differences between the approaches available to cut up the semantic space covered by contrast. Those differences can partly be accounted for by the fact that different studies take different points of departure for the classification of discourse relations (see also Martin 1992: 170–178; Mann & Thompson 1992: 40–42 on this). Some researchers adopt predominantly semantic criteria as a basis for their classification (such as the opposition between contrast, addition, cause/consequence, temporality, etc.), then making more fine-grained distinctions within these broad categories. Another frequent approach consists in focusing on the most prototypical representative(s) of the category, i.e. *but* in English, and *mais* in French, and identifying several subcategories based on their distinct uses (e.g. Lakoff 1971; Anscombre & Ducrot 1977; Csüry 2001). Such approaches generally give rise to classifications in which relations of contrast are grouped together in a single category. On the other hand, some researchers have adopted more specific perspectives for the study of discourse relations, which often lead to a more scattered picture of contrastive relations. Halliday and Matthiessen (2004), for example, start from more 'metaphorical' or functional criteria, i.e. the difference between enhancement, extension and elaboration, only introducing semantic criteria at the second level of their taxonomy. Other studies, such as Martin (1983, 1992), rely on the distinction between internal and external relations as the primary criterion for classification, which has a clear impact on the ways in which the relations are grouped (see also Mann & Thompson 1992: 42 on this). In addition, unlike most accounts, Martin's (1983) classification is based on hypotactic markers instead of markers of conjunction, which also explains part of its distinctiveness. Finally, the originality of Sanders et al's (1992) taxonomy is in large part due to their cognitive approach to meaning, which aims to bring the major discourse relations down to only four cognitive primitives. In conclusion, distinct starting points lead to a situation where equivalent meaning relations end up in different categories, grouped with different meanings.

2.3.2   Contrast in the present book

The previous section attempted to provide an overview of the literature on contrast in English, French and across several languages, while highlighting the great diversity of approaches to the concept. This section aims to specify how contrast is defined in the present book, in the midst of the variety of approaches that exist.

2.3.2.1   *Towards a tripartite view of contrast*

The present study defines contrast as a general semantic category subsuming three subtypes of meaning relations, i.e. opposition, concession and correction. Relations of reformulation, exception and dismissal, on the other hand, are not included in the definition of contrast. Several arguments have motivated the decision to adopt such a definition of contrast. Firstly, the decision to include *only* opposition, concession and correction was prompted by the observation that these relations seem to generate much more consensus than the other three as to their belonging to the category of contrast: while twenty-five, twenty-six and eighteen of the thirty studies reviewed here include opposition, concession and correction, respectively, in the category of contrast, only five studies refer to exception as a subtype of contrast, against four for reformulation, and only three for dismissal. Therefore, it is probably safe to say that if they are in fact contrastive, the latter relations are related to contrast in a much more peripheral way than the former, which seem to constitute the core of contrastive meaning relations. Time and space constraints have led me to limit the present study to the investigation of the more central relations of contrast.

On the other hand, the decision to define contrast as the sum of *all three* of these relations (instead of simply including opposition and concession, for example) stems from the fact that they together summarise the two basic, most fundamental distinctions made in the English and French literature with respect to contrast. Not including one of these seemingly central meanings would certainly have affected the comparability of the study, biasing the definition of contrast in favour of either English (if correction had not been included) or French (if opposition had been excluded). Also keeping cross-linguistic concerns in mind, another argument that has encouraged me to restrict the definition of contrast to the three meanings of opposition, concession and contrast is the fact that a majority of studies aiming to provide a cross-linguistically valid definition of contrast have put forward the tripartite view adopted here, as is clear from Table 3. This suggests that this approach to contrast is well-adapted for a reliable description of meaning relations of contrast across several languages. In addition, adopting the definition used in many other contrastive endeavours will contribute to the comparability of the results obtained in the present study with the results presented in previous contrastive work on contrast.

From a more purely linguistic point of view, there is also lexical evidence in favour of the approach adopted here. On the one hand, in both English and French the three meaning relations of opposition, concession and correction are grouped under a common marker, i.e. *but* and *mais. But* and *mais* can be used to express each of the three meanings, as illustrated in Examples (57) to (62), where (57) and (58) exemplify opposition, (59) and (60) exemplify concession, and (61) and (62) exemplify correction. By contrast, it is much less clear that *but* and *mais* can be used to express either reformulation, exception or dismissal, which all appear to be associated with their own set of more specific markers.[22]

(57)   The cook fried the onions **but** he steamed the cabbage.   (Fraser 1998: 310)

(58)   À Versailles, les sophistications et les rigidités de l'étiquette demeurent toutes-puissantes, **mais** à Paris et dans les grandes villes la noblesse et la bourgeoisie commencent à se mélanger.   (Van de Voorde 1992: 64)

(59)   He is called Siegfried **but** he wears a beret.   (Salkie & Oates 1999: 42)

(60)   Le pain était dur, **mais** elle le mangea.   (Lamiroy & Van Belle 1995: 408)

(61)   The situation is not a new one **but** has been developing over a period of years […].   (Salkie & Oates 1999: 35)

(62)   [Il] n'est pas monté **mais** il est resté à sa place.   (Csüry 2001: 79)

On the other hand, as explained by Izutsu (2008: 652), there is also lexical evidence for viewing opposition, concession and correction as three separate categories in their own right. Although some researchers have argued in favour of merging some of the three meanings (e.g. Grote et al. 1997 view correction as a subtype of concession; Anscombre and Ducrot 1977 view opposition as a mere pragmatic variant of the concessive subtype), this assumption is challenged by the fact that each type of contrast is characterised by a specific set of markers. For example, *by contrast* or *en revanche* are specific to opposition; *even if, although* and *bien que* are exclusively concessive, while *on the contrary* or *instead* for instance, only signal relations of correction. In addition, each of these three relations displays distinctive syntactic behaviours (see Izutsu 2008: 653).

---

22.   Some uses of *but* can express exception, e.g. *He ate all but one cookie*. However, these uses are not viewed as conjunctive markers in the framework of this study, as they occur below clause level. In addition, some researchers have claimed that *but* used as a conjunctive marker can in fact express exception. For example, Sanders et al. (1992: 16) describe the following example as one of exception: *A species can stand a certain amount of hunting, **but** the California condor cannot*. This is the kind of example that I would include in concession, however, as it could be rephrased in the following way: *Although a species can [generally] stand a certain amount of hunting, the California condor cannot.*

Finally, in the present study no explicit distinction is made between internal and external relations of contrast. As explained by Martin (1983: 27), who includes most of the relations of contrast within the broader category of comparison, "external and internal comparative relations are very difficult to distinguish" since "[c]omparative relations are so interpretative in the first place that the distinction is often blurred". This is why, although it is undoubtedly relevant, the internal vs external distinction was only mentioned briefly here but is not taken to be central in the present study.

2.3.2.2   *Features of the relations of contrast*

In this final section on contrast, the three subtypes of opposition, concession and correction are described in some more detail to clearly lay out the main differences between them. As already explained, opposition is a relation which simply states a difference between the contents of the two segments that it links, as in Examples (63) and (64). Both segments are presented as valid but are simply said to exhibit differences. In fact, opposition generally involves a binary opposition (Van de Voorde 1992: 64), in the sense that two pairs of constituents are opposed by the conjunctive marker.[23] In Example (63), for instance, two constituents differ between S1 and S2: the subjects, i.e. 'John' vs 'Sam'; and the complements, i.e. 'tall' vs 'short'. Importantly, the differences signalled by opposition must be stated along a certain area of similarity or common domain which makes the comparison of S1 and S2 relevant. In (63), for example, John and Sam are compared along the common dimension of height. In Example (64), the common ground is the preferred hot drink of the two people compared. By contrast, as Izutsu (2008: 659) explains, an example such as (65), where *rich* and *awake* do not belong to the same semantic domain, makes little sense, unless it is uttered in a highly specific context. A final defining feature of opposition, which differentiates it from the other two types of contrastive relations, is that the two related segments have the same argumentative weight in discourse: neither S1 nor S2 is presented as more important (Mann & Thompson 1992: 39). The relation between S1 and S2 is thus a symmetrical one, which makes it possible to reverse the order of the segments without affecting the meaning of the sequence.

(63)   John is tall. **But** Sam is short.   (Fraser 1998: 310)

(64)   My sister likes coffee, **whereas** my brother prefers tea.

(65)   ?John is rich, **but** Tom is awake.   (Izutsu 2008: 659)

---

23.   Note that in some cases, more than two pairs of elements can be contrasted through opposition, see e.g. *John likes math, Bill likes music, **while** Tom likes chemistry* (Izutsu 2008: 661).

The defining feature of concession is that it necessarily involves implicit informa-tion. In concessive relations, the contrast is set not between two explicit segments, but between an expectation triggered by S1, and the information conveyed in S2 (Salkie & Oates 1999: 33). More specifically, what happens in concession is that two elements that are generally viewed as incompatible are presented as compat-ible: although S1 normally implies a certain event or situation P which conflicts with S2, S2 signals that in this specific case, P does not hold, thus making the apparent incompatibility between S1 and S2 disappear.[24] Therefore, what is in fact rejected in concession is the relation of implication (i.e. the causal relation) that normally holds between S1 and P.[25] In (66), for example, S2 does not deny the fact that it is raining, but the usual consequence of the rain, i.e. that you should stay inside. Two subtypes are generally identified within concession, depending on how much implied information the relation involves (Blakemore 1989: 25–26; Izutsu 2008: 661–667). In direct concession, the pragmatic inference triggered by S1 con-trasts with an explicit message conveyed by S2, as in (66), where the only inference required to understand the relation is that when it is raining, you usually stay inside. In indirect concession, by contrast, inferences must be drawn from both S1 and S2 in order to understand the relation between them, which means that the contrast holds between two implicit messages. In (67), for example, in order to properly understand why S1 and S2 are contrasted, the reader/listener needs not only to draw the inference required in (66), but s/he also needs to understand that S2 suggests that the speaker/writer intends to go out anyway.

(66)   **Although** it's raining I'm going out for a walk.

(67)   It's raining, **but** I have an umbrella.

As was the case with opposition, in concession S1 and S2 are both presented as valid: what is denied is the expectation derived from S1. Unlike opposition, how-ever, concession is an asymmetrical relation: S2 (or what it implies, for indirect concession) is granted more argumentative weight than S1, as the sequence as a whole is meant to reject the implication arising from S1. In concession, S2 is thus presented as having more argumentative force against the implication of S1 than S1 does in favour of it (Anscombre & Ducrot 1977: 28). Finally, unlike opposi-tion, concession does not require S1 and S2 to have something in common. How-ever, for the relation to be coherent the inference drawn from S1 should belong to the same semantic domain as S2 (or what S2 implies, for indirect concession)

---

24.   This explains why concession has also been called 'denial of expectation'.

25.   This is what has led some researchers to classify concession as a causal relation, see Section 2.3.1.

(see Izutsu 2008). Thus, in (66) for example, there is no direct common ground between the fact that it is raining and the fact that the speaker is going out for a walk, but there is between the inference drawn from S1 (i.e. that when it rains, people stay inside), and S2 (which expresses the opposite of staying inside).

In Section 2.2.5, we saw that although generally speaking, conjunctive markers are optional linguistic elements, different discourse relations are not equally likely to be left implicit. It was also pointed out that relations of contrast are among those that frequently need to be signalled explicitly. Interestingly, even within contrast, concession has been demonstrated to necessitate explicit marking more frequently than opposition (see e.g. Asr & Demberg 2013; Grote et al. 1997: 91). The main reason for this is the high degree of complexity of the concessive relation: on the one hand, concession presupposes the presence of implicit information, and therefore demands a certain cognitive effort on the part of the reader to be processed successfully; secondly, concession is complex in the sense that it entails an interaction between the relations of contrast and causality (Kortmann 1991: 160). These elements make concession particularly difficult to process in the absence of an explicit signal, which provides the reader with some instructions for the interpretation of the relation. Thus, within contrast there are also differences in the degree to which the different relations require explicit signalling.

Finally, the relation of correction is characterised by the fact that the first of its segments always serves to deny a certain piece of information, which is either stated explicitly in previous discourse, or can be assumed in the context of utterance of the sentence. S2 then provides the correct alternative to the information rejected by S1. In (68), for example, S1 refutes the idea, probably expressed in previous text, that Harry is happy. S2 then provides the correct information regarding Harry's morale. The negation required by correction is often expressed explicitly (typically by means of *not* in English and *ne…pas* in French), although some instances of correction display a more implicit form of negation, expressed with the help of modal auxiliaries, for example, as in (69) (Fraser 1998: 320). Therefore the specificity of correction is that, as opposed to the other two relations, where the information in both segments is presented as valid, correction fundamentally implies a statement of invalidity of some of the information that it presents (Izutsu 2008: 671; Csüry 2001: 86).[26] As S2 is presented as a substitution of S1, the corrective relation is an asymmetrical one, in which S2 is given more importance than S1 in discourse. Finally, as was the case with opposition, in correction there must be a certain area of similarity between S1 and S2 (Izutsu 2008: 669): one element

---

26.   Technically, though, both S1 and S2 are valid in themselves, and what is presented as invalid corresponds to S1 in its *affirmative* form (see Anscombre & Ducrot 1977).

is common to S1 and S2 (e.g. *Harry* in (68); *she* in (69)), and the two segments express information that belongs to the same semantic domain (e.g. Harry's mood in (68)).

(68)   Harry is <u>not</u> happy. **On the contrary**, he is extremely depressed.

<div align="right">(Fraser 1998: 322)</div>

(69)   She <u>should</u> have taken it. **Instead**, she left it lying there   (ibid.: 320).

The description of the three relations making up the category of contrast has – hopefully – made clear the specificities of each subtype of contrast. In brief, the main distinctive features of each subtype are the following (see Izutsu 2008 for more details): opposition distinguishes itself from concession and correction by being a symmetric relation, with neither of the two segments taking precedence over the other; concession, unlike the other two subtypes, necessarily involves some implicit information, and revolves around the presence of a frustrated expectation. Finally, correction is the only relation of contrast that leads to the rejection of the information expressed in one of its segments. In addition, on the basis of these descriptions it also appears that we can formulate more explicitly what unites these three relations in a single, overarching category (i.e. contrast). Firstly, from this description it appears that all three relations imply a certain degree of common ground between S1 and S2: in all cases, S1 – or what S1 implies, for concession – and S2 must belong to the same semantic domain for the relation to be coherent (see Izutsu 2008 on this). In fact, as Izutsu (ibid.) explains, for a relation of contrast to occur, the information expressed in S1 and S2 should be mutually exclusive within that shared domain: in Example (63), for example, 'John' and 'Sam', and 'tall' and 'short' are mutually exclusive, as you cannot be John and Sam at the same time, nor can you be tall and short at the same time. In concession, the implication of S1 (e.g. staying inside, in Example (66)) and S2 (going out for a walk) are also mutually exclusive. Finally, in Example (68), illustrating the corrective relation, there is a relation of mutual exclusiveness between 'being happy', on the one hand, and 'being depressed' on the other. In a similar vein, some researchers have stated that what is common to the three relations discussed here is that they all seem to imply a negation of some kind: as explained by Csüry (2001: 87), for a contrast to occur it is crucial that a relation of negation should hold between the two segments. Depending on the subtype of contrast, what is denied is either: (i) the similarity of S1 and S2 (opposition); (ii) a relation of implication between S1 and its usual consequence (concession); or (iii) simply the information presented in S1 (correction). This is in line with the Cognitive approach to Coherence Relations put forward by Sanders et al. (1992), in which the contrastive relations are distinguished from the others by being assigned a negative polarity (see Section 2.3.1).

Against this backdrop, the following definition can be formulated for the general category of contrast:

> Contrast is a broad semantic category subsuming the three, more specific, relations of opposition, concession and correction. In contrastive relations, the second segment in the relation instructs the reader to perceive a negation of either the similarity, implication or validity of the information presented in the first segment. In addition, all relations of contrast share the property of asserting the mutual exclusiveness, within a common semantic domain, of the information presented in the segments that they link.

This definition constitutes my *tertium comparationis* for the study of contrast in English and French: all the English and French conjunctive markers used to signal relations that meet this definition will be considered for analysis. In this case, what justifies the comparison of linguistic units across languages is therefore their semantic equivalence.

2.4   Conclusion

The objective of this chapter has been to provide definitions for my two main research constructs, viz. conjunctive markers and contrast. As the present study takes a contrastive approach to CMs of contrast, the chapter started with a brief discussion on the importance of ensuring the cross-linguistic validity of the definitions adopted, through the rigorous establishment of a reliable *tertium comparationis*. Those issues of comparability have underlain all the subsequent discussions. Conjunctive markers were then defined functionally as a set of linguistic items that are used to express logico-semantic relations (such as contrast, causality or addition) between two linguistic units at clause level or above. Although some of the literature pleaded in favour of only including markers linking sentences or larger units in the category of conjunctive markers, preoccupations of cross-linguistic comparability led me to also consider inter-clausal links. Therefore, a broad definition of the category of conjunctive markers was adopted, including the following three types of devices: conjunctive adjuncts, coordinators and subordinators. The last part of the chapter was concerned with the definition of the discourse relation of contrast. Emphasis was laid on the considerable diversity of definitions and approaches to contrast in the literature. Contrast was then defined as a broad semantic category subsuming a set of three relations (i.e. opposition, concession and correction) that shared the property of signalling a relation of mutual exclusiveness between two discourse segments expressing content within the same semantic domain. Negation was also put forward as a defining feature

of contrast, and each of the three relations was shown to be characterised by its own specific way of signalling a negation between the two segments linked. Here again, concerns about the cross-linguistic validity of the study led me to adopt this tripartite view of contrast, obtained by merging the two basic dichotomies made in the English and French literature with respect to contrast. All in all, the present study adopts a semantic-functional *tertium comparationis*: the functional category of conjunctive markers is restricted semantically to only consider conjunctive markers signalling relations of contrast in English and French. In other words, the semantic-functional TC groups together a set of English and French elements that share both the same function (i.e. connecting two clauses or clause complexes) and the same meaning (i.e. contrast).

Corpus-based contrastive approaches to conjunctive markers

This chapter provides a bird's-eye view of the corpus-based contrastive research on conjunctive markers. Section 3.1 provides a general presentation of the field of corpus-based contrastive linguistics, emphasising the benefits that can be gained from a corpus approach to the comparison of languages. Section 3.2 provides a broad overview of the main types of corpus-based contrastive approaches to conjunctive markers in the literature. Finally, Section 3.3 zooms in on some cross-linguistic research on conjunctive markers that is specific to the English-French language pair.

3.1 Corpus-based contrastive linguistics

This section provides a general presentation of the field of corpus-based contrastive linguistics. It starts with a description of the benefits derived from the synergy between contrastive linguistics and corpus linguistics. Section 3.1.2 describes the main types of data used in corpus-based contrastive linguistics, and discusses the advantages and the drawbacks inherent in each type of corpus. Finally, Section 3.1.3 discusses a relatively new trend in corpus-based contrastive linguistics, which consists in assessing the impact of register on the differences between languages.

3.1.1 The benefits of a corpus approach to comparing languages

Contrastive linguistics can be defined as "the systematic comparison of two or more languages, with the aim of describing their similarities and differences" (Johansson 2003: 31). When it emerged in the 1950s, contrastive linguistics (which was then called 'contrastive analysis') had a strongly applied, pedagogical focus: its main purpose was to predict learners' difficulties on the basis of the differences between their mother tongue and the target language (see Lado's 1957 'Contrastive Analysis Hypothesis'; see also Altenberg & Granger 2002: 5

for more details). In the late 1960s, the idea of a direct correspondence between cross-linguistic contrasts and learning problems was largely called into question, and the field fell into disfavour (see Sajavaara 1996 for more details). However, the 1990s saw a strong regain of interest in contrastive linguistics, which went hand in hand with a profound transformation in its goals and methodology and a significantly broadened scope of application – leading Johansson (2012: 46) to speak of "contrastive linguistics in a new key". Today, contrastive linguistics is back to being a flourishing field of research, at the centre of a large volume of publications, conferences and research projects (see Ebeling & Oksefjell Ebeling 2013: Chapter 2 for a detailed history of contrastive linguistics). Among the major factors which have been said to spark the revival of contrastive linguistics is the emergence of computerised corpora, which provided a solid empirical foundation for cross-linguistic comparisons (e.g. Aijmer & Altenberg 1996: 11–12; Aijmer et al. 1996: 73; see Altenberg & Granger 2002 for an overview of the factors which stimulated the revival of contrastive linguistics). This section reviews the ways in which the use of corpora was beneficial for the field of contrastive linguistics.

In present-day linguistics, a corpus is defined as "a collection of (1) *machine-readable* (2) *authentic* texts […] which is (3) *sampled* to be (4) *representative* of a particular language or language variety" (McEnery et al. 2006: 5). In other words, corpora consist of a principled assemblage of authentic texts (i.e. as opposed to invented data based on intuition) that can be processed and analysed (semi-)automatically with the help of computer software. The requirement of representativeness presupposes that the results obtained from a given corpus can be generalised to the language (variety) included in the corpus. The core objective of contrastive linguistics, i.e. the comparison of language systems, typically calls for specific types of corpora, containing data that is representative of the two (or more) languages included in the comparison. These are described in Section 3.1.2. The present section provides a more general account of the ways in which contrastive linguistics can benefit from the use of multilingual corpus data.

One of the first researchers to draw attention to the potential of corpora for the comparison of languages was Filipović who, as early as 1974, stated that "no contrastive project can be regarded as complete before its results are verified and completed by means of a corpus" (Filipović 1974: 62).[1] It was not until the 1990s that this plea was truly heard, with a growing number of initiatives being launched

---

1.   Note that at that time, the term 'corpus' did not specifically refer to machine-readable collections of texts; the type of data advocated by Filipović corresponded to more limited selections of texts, generally stored in print and analysed manually (see Ramón García 2002: 399).

to fully integrate the corpus approach into cross-linguistic comparisons. The main contribution made by corpora in contrastive linguistics has been to considerably increase the reliability and validity of the cross-linguistic comparisons (Johansson 2007: 5; Mikhailov & Cooper 2016: 15). Prior to the emergence of corpora, contrastive linguists customarily relied on their intuition as the main source of information on the differences between the languages compared (Granger 2003: 18). Alternatively, contrastivists formulated very general differences between languages based on very limited sets of examples extracted from equally limited samples of linguistic data (e.g. a few texts in language A and their translations into language B). A first problem with these early contrastive methods is that, as pointed out by Granger (1996: 40) "intuitions are rarely wrong, but […] they often give only a partial – and hence misleading – view of reality". Studies based on small collections of authentic texts in the languages compared – generally in the form of original texts together with their translations – arguably have more reliable grounds for the identification of similarities and differences between languages. However, the amount of data in these studies is far too limited to allow for generalisations about the differences between the languages.

The shortcomings characterising this early generation of contrastive studies were largely overcome by the corpus methodology. The large quantity of data made available through corpora, along with the (partial) automation of the analyses – which made possible the analysis of substantially increased volumes of data in roughly equivalent amounts of time – allowed for much more reliable generalisations than was the case with the methods described above. In addition, the quantifications permitted by corpus analyses make it possible to complement statements on differences in terms of the structural possibilities offered by two (or more) language systems, with statements on diverging preferences or 'optimal choices' within linguistic systems offering similar structural possibilities (Schmied 2009: 1143). Moreover, by granting access to extensive amounts of data from a variety of language registers, the corpus approach made it possible to balance contrastive claims according to the communicative situation, a factor which has recently been demonstrated to be of great significance for the comparison of languages (see Section 3.1.3 for more details). In summary, the emergence of corpora in contrastive linguistics came as a particularly appropriate response to a "steadily growing realization that cross-linguistic studies [could not] rely on introspection or scanty empirical evidence, but must be firmly based on naturally occurring language used in a variety of situations" (Aijmer & Altenberg 1996: 11).

The benefits of reliability and representativeness described above, while providing an undeniable boost to contrastive linguistics, are not specific to cross-linguistic analyses, but are valid for all types of linguistic research resorting to corpus data. In addition to these advantages, Aijmer and Altenberg (1996: 12) list

four possibilities afforded by corpora that are fully specific to multilingual data (see also Johansson 2003 for a more detailed discussion of these four possibilities):

i.   they give new insights into the languages compared – insights that are likely to be unnoticed in studies of monolingual corpora;

ii.   they can be used for a range of comparative purposes and increase our knowledge of language-specific, typological and cultural differences, as well as universal features;

iii. they illuminate differences between source texts and translations, and between native and non-native texts;

iv.   they can be used for a number of applications, e.g. in lexicography, language teaching and translation.

The first two possibilities mentioned by Aijmer and Altenberg make clear the wide array of objectives pursued by corpus-based contrastive linguistics. The aims of the comparisons vary from rather general – e.g. trying to identify elements that are common to several languages (see e.g. König 2012; van der Auwera 2012 on the contribution that contrastive linguistics can make to typology and the relationship between these two fields) – to very particular, where the main objective is to learn more about one language by studying it through the prism of another language. The contrastive approach can indeed bring out features of an individual language that would be very hard to detect if that language was investigated in isolation. Altenberg (2002) provides a good example of the way in which multilingual data can inform our understanding of the individual languages compared. Through examining the English translations of the Swedish concessive marker *ändå*, Altenberg was able to identify five main senses of the marker, viz. basic concession, dismissal, explanation, qualification and hypothetical wish. Similarly, Beeching (2013) resorts to translation data to investigate the semantic change undergone by the pragmatic marker *quand même*. By looking at the English translations of *quand même*, she finds evidence that the marker has developed an interpersonal meaning (which she calls 'relational') in addition to its initial contrastive meaning.

The third of the possibilities listed by Aijmer and Altenberg – i.e. the fact that multilingual corpora shed light on differences between source texts and translations, and between native and non-native texts – alludes to two kinds of welcome rapprochements that were brought about by the incorporation of corpus methods into contrastive linguistics: (i) the rapprochement between contrastive linguistics and translation studies; and (ii) the rapprochement between contrastive linguistics and the study of learner language. Today, contrastive linguistics and translation studies are considered to be distinct fields of academic research, characterised by

different goals and objects of study: while the objective of contrastive linguistics is to uncover the similarities and differences between languages, translation studies rather aim to uncover the mechanisms underlying the translation process (Aijmer 2008: 286). Despite these differences, however, the two fields also have many elements in common, among which the concurrent study of several languages, and the importance attached to translations. Another major area of common ground between contrastive linguistics and translation studies lies in the use of multilingual corpora: in the same way as contrastive linguistics, in the 1990s the field of translation studies was profoundly transformed by the emergence of the corpus methodology, advocated mainly by Baker as an alternative to the introspective methods that had hitherto prevailed (see e.g. Baker 1993; 1995). The integration of corpus methods thus gave contrastive linguistics and translation studies a common concern in the building of multilingual corpora, and the development of appropriate tools and methods for their analysis (Vandepitte & De Sutter 2013: 37; Granger et al. 2003: 9). This led to an increased cooperation between the two fields, which greatly profited both contrastive linguistics and translation studies (see e.g. Granger 2003; Ramón García 2002; Vandepitte & De Sutter 2013 for more on this). Although the two disciplines differ in their ultimate goals, they each shed light on a specific aspect of the same phenomenon (i.e. how a given meaning in language A is rendered in language B). This makes their findings highly relevant for the other discipline and entails that a lack of communication between contrastive linguistics and translation studies can be detrimental to both fields. As Granger (2003: 25) puts it:

> [L]ack of familiarity with TS [translation studies] findings may lead CL [contrastive linguistics] researchers to interpret their data in terms of differences between language systems when they result from translation norms or strategies, while TS researchers may similarly misinterpret their data because of a lack of awareness of a systematic difference between the two language systems established by CL.

By providing a fertile ground for greater cooperation between contrastive linguistics and translation studies, corpora have thus made a valuable contribution to both fields. This fruitful rapprochement is crystallised in a number of academic undertakings situated at the interface between translation studies and contrastive linguistics, among which the *Using Corpora in Contrastive and Translation Studies* (UCCTS) conference series – and the volumes that have followed from these conferences (see e.g. Xiao 2010b; Granger & Lefer 2020).

A second important rapprochement made possible by the development of corpora was the one between contrastive linguistics and interlanguage analysis. This rapprochement is epitomised by Granger's (1996) Integrated Contrastive Model, which advocates the combined comparison of (i) native and learner data in a given

language; and (ii) data in the learners' mother tongue and in the target language. This design enables the researcher to either predict or interpret tendencies observed in the learner's interlanguage by reference to differences between the target language and his/her mother tongue.[2] Conversely, instances of overuse, underuse or misuse in learner data can provide an impetus for contrastive linguistics (Hasselgård 2010a: 101). The difference with Lado's (1957) Contrastive Analysis Hypothesis is that the Integrated Contrastive Model always implies some degree of to-ing-and-fro-ing between the two types of comparison: the predictions formulated on the basis of contrastive analyses are then checked against learner data; similarly, a given difference between native and learner language can only be attributed to mother tongue influence if it is in fact supported by differences between the learner's mother tongue and the target language (Granger 1996: 46). In a more recent article, Granger (2018) even demonstrates the merits of a combination of data and methods from contrastive linguistics, translation studies *and* learner corpus research. She shows that such an integrated approach makes it possible to successfully tease apart genuine systemic differences between languages and differences stemming from the translation process (see Section 3.1.2 for more on this issue). All in all, the emergence of corpora thus gave rise to multiple areas of fruitful cross-fertilisation between contrastive linguistics and other fields of research.

Finally, as underlined by Aijmer and Altenberg (1996: 12), the emergence of multilingual corpora has been instrumental for the development of new applications of contrastive linguistics. As already explained earlier, while the initial application of contrastive analysis – i.e. to predict learners' difficulties based on the differences between languages – was largely thwarted in the 1960s, the appearance of corpora gave birth to a range of new, probably more realistic applications of cross-linguistic research. For example, the emergence of corpus data in contrastive linguistics offers great potential to enrich bilingual dictionaries. Multilingual corpora – and in particular those containing original texts along with their translations; see Section 3.1.2 – constitute an invaluable repository of alternative equivalents to the "artificial or invented equivalents that are so often found in 'old-fashioned' dictionaries" (Mikhailov & Cooper 2016: 150). The value of multilingual corpora for lexicography has been discussed by Salkie (2008), for example, who demonstrates that a wide range of translation equivalents can be accessed through corpora, which also provide the necessary quantitative information to help lexicographers decide what equivalents are worth including in a dictionary,

---

2.   A number of additional comparisons are also possible (e.g. learner vs translated language; comparisons across learner populations). It falls out of the scope of this study to expand on these issues.

and what equivalents are too context-bound to deserve mention. Granger and Lefer (2013), on the other hand, have stressed the great value of multilingual corpora to improve the translations of multi-word units provided in bilingual dictionaries, taking the example of phraseological units containing English *yet* and French *encore* (see also Granger in press on the usefulness of multilingual corpora to improve the phraseological coverage of bilingual dictionaries). Other examples of applications of corpus-based contrastive research include: (i) foreign language teaching, through the development of corpus-based pedagogical material and teaching methodologies; (ii) translator training, notably with the help of data-driven learning, which consists in presenting novice translators with multilingual corpus data, in order to draw their attention to differences in behaviour between the languages studied, and/or to stimulate their reflection on the ways in which translators have addressed specific problems of equivalence in the corpus (see e.g. Bowker 1999; 2003); (iii) natural language processing, and more particularly the improvement of machine translation and terminology extraction (see e.g. Granger 2010: 7–9; Mikhailov & Cooper 2016: chap. 6 for more details).

All this being said, it should be noted that the corpus methodology also poses a number of problems for linguistic analysis. The main issue is that corpus analysis mainly allows for form-based searches (Granger 2003: 23). For research focusing on a specific linguistic form (e.g. the suffix *–ish*; the adjective *good* or the complementation of the verb *take*), the corpus methodology is ideal, as a number of software tools enable researchers to simply retrieve all the occurrences of a given form automatically from the corpus. By contrast, for research objects that cannot be reduced to a well-defined (set of) form(s) (e.g. metaphors, idioms, or most semantic categories), the use of corpora can prove much more problematic (Mikhailov & Cooper 2016: 10–11). There are a number of solutions around this difficulty, among which the compilation of a list of forms expressing the meaning or function under investigation. However, such search lists, even if carefully designed, will rarely be fully exhaustive. Thus, in these cases, it is very likely that resorting to corpus methods will lead the researcher to miss some relevant instances, which the human mind would have picked up by carrying out a manual analysis. Likewise, the fully form-based type of analysis afforded by corpus linguistics misses all the relevant forms that would be affected by the encoding problems, typographical errors, etc. that inevitably occur in the corpus compilation process and which, once again, a human analyst would have been capable to trace back to the initial form (ibid.: 11–12). Despite these problems, the numerous advantages offered by corpora are generally considered to far outweigh their shortcomings – not to mention that the human mind is far from flawless and can be prone to oversights that do not crop up with automatic corpus analyses. Yet other problems encountered by users of multilingual data stem from the fact that

the design and compilation of these corpora, in particular, present a number of challenges, as discussed below.

3.1.2   Types of corpora in contrastive linguistics

Two main types of corpora are commonly used in contrastive linguistics, namely comparable corpora and translation corpora.[3] Comparable corpora "consist of original texts in each language, matched by such criteria as time of composition, domain, genre, intended audience, etc." (Johansson 1998: 5). Translation corpora, on the other hand, are made up of original texts in one language, together with their translations into one (or several) other language(s). Translation corpora can be either unidirectional, i.e. contain texts in a single translation direction, or bidirectional, i.e. contain original texts in language A and their translations into language B, and vice-versa. Each type of corpus has both advantages and limitations. Interestingly, the strengths of one type of data generally correspond to the limitations of the other type, which has led many researchers to advocate a combination of the two types of resources, so that the shortcomings of one corpus type can be compensated by the strengths of the other (see below). As corpus analysis can be a fairly time-consuming undertaking, however, such an approach is not always possible. Therefore, it is useful to be aware of the features of each corpus type, so as to determine which one is best-suited for a specific research purpose. Essentially, the advantages and limitations of each type of multilingual corpus can be described along four main dimensions, viz. (i) accessibility, (ii) text-type comparability, (iii) equivalence of the linguistic units compared and (iv) reliability of the data (Granger 2010: 17).

   One element that may encourage contrastivists to prefer comparable corpora to translation corpora pertains to the greater availability of comparable as compared to translation data. Because they contain original data in the two (or more) languages compared, comparable corpora are much easier to collect than their translation counterparts. In fact, in many cases it is even possible to assemble them by grouping already existing corpora originally compiled for monolingual

---

3.   Different terms have been used to refer to the various types of multilingual corpora. The term 'parallel corpus', in particular, has been used to refer to both comparable data (see e.g. Aijmer & Altenberg 1996; Lauridsen 1996) and translation data (see e.g. Teubert 1996; Aijmer 2008; McEnery & Xiao 2007). In addition, the terms 'comparable corpus' and 'translation corpus' have received different definitions than the ones provided here (see e.g. Granger 2003: 19–20; McEnery & Xiao 2007: 18 for a discussion of issues of terminology with respect to multilingual corpora). This study adopts the terminology suggested by Granger (2003) and Johansson (2007). In view of its ambiguity, the term 'parallel corpus' is avoided altogether.

research (Aijmer 2008: 277). This is the method adopted by Crible (2018), for example, who assembled a comparable corpus of speech across eight registers by matching the spoken subcorpora included in the British component of the International Corpus of English (ICE-GB) with available corpora from corresponding registers in French. By contrast, it can be very difficult to get access to translation data that is fully in accordance with the purposes of the cross-linguistic study. Firstly, because not all types of linguistic activities are translated, translation data is generally restricted to a limited number of – typically written – text types (see e.g. Johansson 1998: 6; Mauranen 1999). The type of data available in translated form typically includes best-selling fiction, along with a range of highly specialised texts such as EU documents or instruction manuals, which may call into question the generalisability of the findings that their analysis gives rise to. In addition, the amount of data available is highly dependent on the language pair as well as the translation direction investigated (Altenberg & Granger 2002: 9). For example, Johansson (2007: 13) explains that one of the problems that he encountered when building a bidirectional English-Norwegian translation corpus was that the range of translated texts was far greater in the English-Norwegian than in the Norwegian-English translation direction. Another element impeding the compilation of large, good-quality translation corpora is the fact that, for certain text types, it is very difficult to identify which of the two texts is the original and which is the translation, although this information is of course crucial for cross-linguistic analyses (Lauridsen 1996: 66). Finally, translation corpus alignment, which consists in pairing units in the original texts with their translations at paragraph, sentence or even sometimes word level, constitutes a major challenge for anyone embarking on the task of compiling a translation corpus. For all these reasons, translation corpora – when they are in fact available – are generally imbalanced and fairly limited in size, which severely limits their potential for contrastive analysis, especially for the study of low-frequency phenomena (Altenberg & Granger 2002: 9).

While the compilation of comparable corpora is facilitated by the wide availability of original texts across various language registers, it also poses a major challenge, i.e. the cross-linguistic comparability of the texts included in the corpus. In order to make sure that the differences identified reflect genuine contrasts between languages, and are not due to discrepancies in terms of register, time of production, target audience, etc. between the texts included in the corpus, it is essential to ensure the maximum comparability of the texts making up the comparable corpus (e.g. Johansson 2007: 10; Connor & Moreno 2005). Ensuring the comparability of texts across languages is far from a straightforward matter, notably because some text types are culture-specific and do not have any equivalent in other languages (Granger 2010: 18). In addition, as pointed out by Neumann (2014: 42), "seemingly comparable registers may still display differences in their actual realization".

Thus, for example, the fact that two sets of texts both represent the register of 'book reviews' or 'editorials' in their respective language does not necessarily ensure that the conventions of that register are equivalent in the two languages, and therefore that these texts are in fact fully comparable (see Chapter 5 for more on this). These complex questions of data comparability are irrelevant for translation corpora, since the texts that they contain, being translations of each other, are in principle equivalent from a semantic, pragmatic and textual point of view (McEnery & Xiao 2007: 19; Granger 2010: 18).

Another problem with comparable data is that it is sometimes difficult to determine what linguistic units to compare across languages or, in other words, what *tertium comparationis* to adopt for the cross-linguistic comparison (see Chapter 2 for a definition of *tertium comparationis*). With translation data, it is fairly easy to know what to compare, i.e. to relate expressions that have the same meaning and function across languages, since the equivalents of a given form in language A are directly accessible through their translations in language B (Johansson 1998: 5). In fact, translation has been described by James (1980: 178) as "the best available *tertium comparationis*". Comparable corpora, on the other hand, do not provide such a ready-made *tertium comparationis*, available by screening automatically-extracted occurrences of linguistic items aligned with their translations. The choice to resort to comparable data will therefore entail an upstream effort of reflection in order to ensure that a language-neutral basis for comparison of the languages is adopted. As was clear from our discussions in Chapter 2, this can be a truly challenging task, which also largely determines the validity of the subsequent cross-linguistic comparisons.

A final element to be taken into consideration when selecting a corpus for contrastive analysis pertains to the authenticity and reliability of the data. Comparable corpora contain original texts in each of the languages compared, representing ordinary and spontaneous language use, generally produced by native speakers, and unaffected by the influence of other languages (Granger 2010: 19).[4] As such, they allow for safe conclusions to be drawn on the similarities and differences between the languages compared, notably in terms of frequencies of use of linguistic phenomena (Johansson 2007: 10). Translation, on the other hand, has been said to give a distorted picture of the language that it represents, which may differ in a number of ways from what a native speaker would have produced spontaneously (see e.g. Teubert 1996; Lauridsen 1996). More precisely, two main elements have been said to set translated language apart from original language. Firstly,

---

4.   Although the internationalisation of society has made most languages permeable to the influence of other languages, even in their original form.

translated language has been shown to contain traces of influence, or "unmistakeable fingerprints", from the source language, referred to as translationese (Gellerstam 1996: 54). In other words, translators, even if they are highly proficient in both the source and the target languages, tend to make choices that are influenced by the features of the source text, in a way that is "strong enough to make L2 data perceptibly different from the target L1" (McEnery & Xiao 2007: 22). The literature provides ample evidence of source language influence in translation. For example, studies have reported excellent results (around 90% accuracy) in (i) automatically distinguishing between original and translated texts in the same language and (ii) automatically identifying the source language of a given translated text (see Granger 2018; Kurokawa et al. 2009). Another feature of translations that has been said to mark them out from original language is that they are influenced by so-called 'translation universals', i.e. "features which typically occur in translated texts rather than original utterances and which are not the result of interference from specific language systems" (Baker 1993: 243) – thereby being traceable to the translation process itself. One such feature, which has been discussed at length in the literature, is the tendency of translated texts to be more explicit than their original equivalents, including with respect to cohesion (see Chapter 5 for more details). Other translation-induced phenomena include normalisation (defined as a greater tendency to conform to the norms of the target language, sometimes to the point of exaggeration) and simplification (e.g. in terms of lexis or grammatical structures; see e.g. Laviosa 2009 for more information on translation universals).[5]

These inherent features of translated texts have prompted some researchers to seriously call into question the reliability of translation data for contrastive linguistics. One of the most categorical rejections of translation corpora is voiced by Teubert (1996: 247):

> Translations, however good and near-perfect they may be (but rarely are), cannot but give a distorted picture of the language they represent. Linguists should never rely on translations when they are describing a language. That is why translations have no place in reference corpora. Rather than representing the language they are written in, they give a mirror image of the source language.

In a similar vein, Lauridsen (1996: 67) advocates the use of comparable corpora for contrastive linguistics, claiming that "one should refrain from using translation corpora unless the purpose of the linguistic analysis is either to evaluate the

---

5. These features of translations have led some researchers to refer to it as a 'third code', i.e. a hybrid system, which displays features of both the source and target languages while not being truly equivalent to either of them. In this sense, translation can be described as a kind of interlanguage (Castagnoli 2009: 8ff; Granger 2018).

translation process or to criticize the translation product on the basis of a given translation theory". Admittedly, such phenomena as translationese and translation universals, if not controlled for, can have the consequence of formulating cross-linguistic contrasts that are in fact attributable to the translation process, or missing differences that were masked by the influence of the source language.

Nevertheless, a number of researchers have held more moderate views on translation data, also stressing their great potential for contrastive linguistics (e.g. Aijmer 2008: 280; Ebeling & Oksefjell Ebeling 2013: 44). Mauranen (1999: 182), for example, states that translation corpora are "indispensable for contrastive language study". As she explains, they are unique resources for the following tasks: to uncover the extent of the equivalence between two expressions across languages, to help us capture the meaning of linguistic units by giving us access to their translation paradigms, to be used as a tool for word sense disambiguation, or even simply to invite further research based on original data in the languages compared. Thus, instead of rejecting translation data point-blank, some researchers have suggested solutions to circumvent its limitations. One solution is to resort to bidirectional translation data: by verifying whether or not the differences observed are reflected in both translation directions, it is possible to determine whether they stem from genuine cross-linguistic differences, or are effects of the translation process (McEnery & Xiao 2007: 22). Another, widely accepted solution is to combine translation and comparable data. In this section, the advantages and limitations of these two types of corpora were described along four dimensions, i.e. availability, text-type comparability, equivalence of the linguistic units compared and access to fully original, unmediated language (cf. Granger 2010). A summary of this discussion is provided in Table 4. As is clear from this table, the two types of data are complementary, the strengths of one corpus type corresponding to the weaknesses of the other.

Table 4. Advantages and limitations of comparable and translation corpora (based on Granger 2010: 17–18)

	Translation corpora	Comparable corpora
Availability	✘	✓
Text-type comparability	✓	✘
Equivalence of the linguistic units	✓	✘
Original data	✘	✓

The combination of the strengths of comparable and translation data is what the 'bidirectional translation model', put forward by Johansson and his team

(see e.g. Johansson & Hofland 1994; Johansson 2007) for the contrastive analysis of English and Norwegian, aims to achieve. The English-Norwegian Parallel Corpus (ENPC) consists of "a comparable corpus and a [bidirectional] translation corpus […] combined within the same overall framework [where] each type can be used to control and supplement the other" (Johansson 2007: 11).[6] The model, which is represented in Figure 6, allows for a multiplicity of comparisons, "mak[ing] it possible to distinguish between language differences and translation effects" (ibid.: 12). It permits comparisons between: (i) original texts across languages – i.e. it can be used as a comparable corpus ; (ii) original texts and their translations into the other language – i.e. it can be used as a translation corpus; (iii) original and translated texts in the same language (which makes it possible to identify what features were induced by the translation process); and (iv) translated texts across languages (a type of comparison that is mainly used in translation studies, but can also greatly benefit contrastive linguistics, see e.g. Granger 2018).[7]
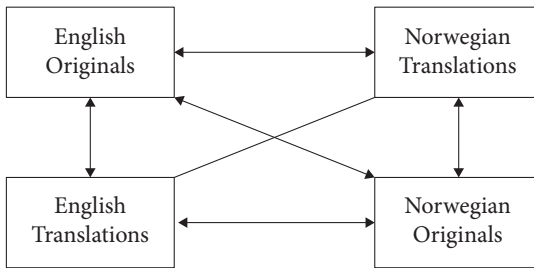
Figure 6. The ENPC model (Johansson 2007: 11)

In Johansson's model, the combination of comparable and translation data can occur in either of the following ways: (i) identifying similarities and differences

---

6. This model has also been used for the compilation of other multilingual corpora, such as the English-Swedish Parallel Corpus (Aijmer & Altenberg 2000) or the Poitiers-Louvain Échange de Corpus Informatisés (PLECI), an English-French bidirectional translation corpus compiled in collaboration between the Centre for English Corpus Linguistics (Université catholique de Louvain) and the FoReLL research laboratory (Université de Poitiers).

7. Note that such a model can be further extended by adding other languages into the picture (e.g. English and Norwegian original texts translated into German, and vice-versa): as explained by Johansson (2007: 18), the inclusion of more languages will allow researchers to identify more clearly the features of each language system, but also the features of translation (see also Granger 2018; van der Auwera 2012 on the benefits of involving more than two languages in cross-linguistic comparisons).

in the translation corpus, then checking for translation effects with the help of comparable data; or (ii) start by describing and comparing patterns in each language based on the comparable data, and then use the translation corpus to refine the results by identifying translational correspondences (Johansson 2007: 33). The approach chosen depends on the topic and objectives of the study (ibid.). In sum, such an integrated model allows contrastive researchers to exploit multilingual corpus data to the full of its potential.

3.1.3   Register-sensitive contrastive linguistics: An emerging trend in contrastive linguistics

As explained in Section 3.1.1, one of the potentialities offered by the use of corpora in contrastive research has been to provide access to data that is representative of a range of registers, i.e. varieties of language as determined by the situational context (Halliday 1978: 32; see Chapter 4 for more information on register in SFL). Stylistically-varied multilingual corpora make it possible to refine claims on the similarities and differences between the languages compared by reference to the communicative situation, whose influence on the frequency and patterning of linguistic phenomena has long been demonstrated in the monolingual field. Yet, for many years, this potential of corpora was not fully exploited by researchers in contrastive linguistics. For a long time, contrastive linguists considered the languages compared as monolithic entities, making claims about the differences and similarities between them as though they were valid in all situations. In other words, as Neumann (2012: 193) explains, traditional contrastive studies have tended to be system-based, focusing primarily on the comparison of the possibilities offered by the language systems, in spite of the fact that "there is a well known difference between the potential provided by language systems and the choices made by speakers of a given language, particularly in terms of the more specific options filtered by different registers".

     In recent years, however, contrastive linguists have grown increasingly aware of the necessity to adopt a register-sensitive approach to the comparison of languages (see e.g. the edited volume by Lefer & Vogeleer 2014; see also Aijmer & Lewis 2017), and have gradually started answering Johansson's (2007: 304) plea to "extend contrastive studies by taking into account the variation across registers within languages".[8] A number of studies have raised awareness to the fact that

---

8.   It should be noted that a similar interest for register variation is observed in translation studies, where it has received more focus than in contrastive linguistics from the start (see e.g. Neumann 2012; Kruger & van Rooy 2012; Delaere & De Sutter 2017 on the influence of register on the translation process).

cross-linguistic differences are not always stable across communicative situations. In morphology, for example, Lefer and Grabar (2015) have investigated the French translations of English evaluative prefixes across four registers, viz. parliamentary debates, TED Talks, fiction and news. They have uncovered marked differences between TED Talks and the other three registers, with TED Talks displaying a significantly higher proportion of prefix-by-prefix (i.e. literal) translations and a much lower frequency of periphrastic translations than the other three registers.[9] In the field of phraseology, Granger (2014) has compared the use of English and French lexical bundles in a comparable corpus of newspaper editorials and parliamentary debates, with a view to testing the hypothesis that French tends to be more formulaic than English. Her results show that this hypothesis is only confirmed for the editorial register, whereas in parliamentary debates the frequency of bundles is much higher in English than in French. These results thus suggest that differences between languages may be register-dependent. By contrast, register-sensitive contrastive studies can also uncover cross-linguistic differences that are consistent across communicative situations. Kunz and Lapshinova-Koltunski (2015), for example, compare the use of cohesive devices in English and German based on a corpus made up of ten written and spoken registers. They report a certain stability of the cross-linguistic differences in the use of cohesion across registers, and find that the differences are more pronounced between languages than between registers (see Section 3.2.2 for more details).

   In this context of growing concern for register variation in contrastive linguistics, we have witnessed the emergence of multilingual corpora including data from a variety of registers, such as the CroCo corpus (which contains original and translated texts in English and German across eight registers, see e.g. Hansen-Schirra et al. 2012) or the Dutch Parallel Corpus (including data across five registers for the Dutch-English and the Dutch-French language pairs, see e.g. Macken et al. 2011). In parallel, Neumann (2014: 42–43) has designed a typology identifying three main types of register-sensitive cross-linguistic studies: (i) studies that control the corpus under investigation for register by including texts from one register only, thus acknowledging register to be an influential factor in cross-linguistic comparisons; (ii) studies that compare specific linguistic phenomena across several registers; and (iii) studies that focus on cross-linguistic differences

---

9.   Note that Lefer and Grabar actually take the reflection further by also pointing out that other factors differentiating multilingual corpora, such as the translation mode (e.g. translation vs subtitling) or the expertise of the translator may have an impact on the type of equivalences emerging from translation data.

between registers themselves instead of differences in the use of a given linguistic phenomenon.

In accordance with the recent set of studies seeking to integrate register variation into contrastive descriptions, and in view of the fact that the impact of register on the use of conjunctive markers has been conclusively demonstrated in the monolingual field (e.g. Smith & Frawley 1983; Liu 2008; see Chapter 7 for more details), part of the analyses presented in this book will assess the impact of register variation on the differences between English and French uses of CMs of contrast. Such an approach is thus in line with the second type of register-sensitive contrastive studies identified by Neumann. This is another aspect of this study for which the choice of the SFL framework proves particularly well-suited, since SFL has been described as an ideal framework for the cross-register comparison of languages (Teich 2003: 28; Neumann 2010: 88).

3.2   Corpus-based contrastive research on conjunctive markers

Extensive cross-linguistic research has been carried out on conjunctive markers: these items have been the focus of a large number of studies in contrastive linguistics, but also in translation studies – where they have notably been investigated in relation to the translation universal of explicitation (e.g. Becher 2011; Espunya 2007; Károly 2017) – and learner corpus research – which has compared native and learner usage in order to identify patterns of overuse, underuse and misuse of CMs across a number of learner populations (e.g. Granger & Tyson 1996; Altenberg & Tapper 1998; Field & Yip 1992). This section provides an overview of cross-linguistic research on conjunctive markers. Due to space constraints, it nevertheless focuses more specifically on (i) corpus-based (ii) contrastive research (iii) on written language, which is the most directly relevant type of research for the present study. Admittedly, other types of cross-linguistic studies can also provide valuable insights into the differences between languages in terms of CM use. For example, research in translation studies sometimes refers to differences between the language systems to disentangle effects of the translation process from source-language influence. Research in other fields can therefore be very useful for the interpretation of the corpus results presented in Chapters 7 and 8 and will be referred to whenever relevant in the empirical part of this study.

Two main types of contrastive studies have compared the use of conjunctive markers in written corpora. A first group of studies has been concerned with assessing the degree of equivalence between CMs across languages. These studies, which tend to focus on a limited set of markers that they describe in extensive detail, are presented in Section 3.2.1. A second set of studies have adopted a more

onomasiological approach to the cross-linguistic comparison of CMs: they have attempted to investigate how a given meaning or function tends to be expressed in two or more languages, instead of focusing on a closed set of forms. These studies are the object of Section 3.2.2.

3.2.1 Cross-linguistic equivalences between conjunctive markers

One very widespread type of corpus-based contrastive approach to conjunctive markers aims to assess the degree of equivalence between CMs across languages. Such research most typically takes the form of case studies, which concentrate on one or two pairs of CMs – usually cognates or semantic-pragmatic equivalents (Aijmer & Simon-Vandenbergen 2011: 234–235). Examples include Degand's (2004) study on French *puisque* vs Dutch *aangezien*; Mortier and Degand's (2009) comparison of French *en fait* and Dutch *eigenlijk*; Lewis's (2006) analysis of the pairs *au contraire/on the contrary* and *on the other hand/en revanche* in French and English or Buysse's (2017) comparison of English *so* and Dutch *dus* (see also Lamiroy & Vanderbauwhede 2016; Hasselgård 2014a; Nome & Hobæk Haff 2011; Fretheim & Johansson 2002; Lewis 2005 for similar studies). In addition, a few studies are rather broader in scope, and attempt to uncover the degree of correspondence between the various members of semantically-coherent categories of markers – e.g. English and French backward causal markers (Zufferey & Cartoni 2012), French and Dutch backward causal markers (Degand & Pander Maat 2003) or English and Swedish concessive and resultative conjunctive adjuncts (Altenberg 2002; 2007, respectively).

In these studies, the researchers generally rely on translation data to identify the equivalences between CMs, resorting to measures such as 'mutual correspondence', which calculates the frequency with which two CMs are translated by each other in a bidirectional translation corpus (Altenberg 1999: 254). One major finding that has emerged from such research is that equivalences between CMs across languages are usually complex and only partial, even between typologically-close languages. Thus, for example, scores of mutual correspondence obtained on the basis of translation corpora are generally low, even between CMs which are generally perceived as clear equivalents, and described as such in bilingual dictionaries. In his translation-based analysis of resultative CMs in English and Swedish, for instance, Altenberg (2007) reports that the members of the three pairs of cognates *so/så, therefore/därför* and *then/då*, although closely related in meaning, are translated by each other in just over half of the cases. In addition, CMs rarely display a single, clear equivalent but typically have fairly varied ranges of translation equivalents (see e.g. Hasselgård 2014a on the English and French translations of Norwegian *dessuten*; Fretheim & Johansson 2002 on the English correspondents to Norwegian *likevel*).

Based on these observations, researchers have attempted to uncover the reasons for the lack of cross-linguistic equivalence between CMs, and pinpoint the factors allowing to predict the most appropriate equivalent(s) of a given CM in a specific context. The lack of equivalence between CMs across languages has often been explained by reference to fine-grained semantic differences between them, including, for instance, divergences in the degree of subjectivity of the relations that they signal, differences in the information status – given vs new – of the segments that they link (e.g. Zufferey & Cartoni 2012), or in the range of concessive nuances that they can encode (e.g. Altenberg 2002; Fretheim & Johansson 2002). Patterns of correspondence between CMs have also been accounted for by reference to stylistic or syntactic factors. For example, based on an English-Dutch bidirectional translation corpus containing texts from five written registers, Buysse (2017: 59) shows that *so* is translated by *dus* more frequently than the opposite – a phenomenon referred to by Altenberg (2007: 258) as 'translation bias'. He explains that this translation bias is especially noticeable in the more formal registers included in the corpus, due to a relatively frequent correspondence between Dutch *dus* and English *therefore* – i.e. a more formal variant of English *so* – in these text types. Similar results are also obtained by Hasselgård (2014a) and Altenberg (1999: 260), who demonstrate that the preferred equivalent(s) of a given marker are not always the same across text types (see also Dupont & Zufferey 2017 on the impact of register on the translation of CMs). With respect to the influence of syntax, Hasselgård (2014a) and Fretheim & Johansson (2002) observe that the translation equivalents chosen for the Norwegian adverbial markers *dessuten* and *likevel* – respectively – vary in function of the position that these markers occupy in the sentence.

Translation-based studies of CM equivalences thus typically provide very detailed descriptions of the markers investigated, thereby increasing our understanding of the differences and commonalities between individual conjunctive markers, not only across languages, but also within language systems. As explained in Section 3.1.1, one of the many advantages of multilingual data is that it provides unique insights into the individual languages compared – and not simply the differences between them. Through the establishment of translation paradigms, these studies increase our understanding of the fine-grained usage differences that exist between CMs expressing the same type of relations (e.g. causality, contrast) within a language, thereby helping to explain why CMs are rarely interchangeable (Zufferey & Cartoni 2012: 233; see also Zufferey 2012). For example, Nome and Hobæk Haff's (2011) analysis of the Norwegian translations of French *donc* increases our understanding of the various usage patterns of this marker, with the distinct Swedish equivalents reflecting different – although closely related – semantic nuances of *donc*. Likewise, Fretheim and Johansson (2002) gain valuable

insights into the semantic and pragmatic features of the Norwegian marker *likevel* through the analysis of its translations into English. They explain that *likevel* does not correspond to a single form in English, but rather displays "a range of correspondences which reveal various facets of its use" (ibid.: 277). All in all, these studies provide much support in favour of the claim that translation corpora are "unrivalled when it comes to discovering the translation correspondences and displaying them as a set or paradigm" (Aijmer 2008: 278), in a way that provides rich information on the meaning and functions of the linguistic items investigated (ibid.: 284).

Another phenomenon that has been shown to partly explain the generally low degrees of mutual correspondence between CMs pertains to cases of so-called 'zero correspondences', i.e. cases where a CM in the source text has no counterpart in the target text (i.e. omissions), or cases where a marker in the target text has no source-text equivalent (i.e. additions; see Johansson 2007: 26). In Dupont and Zufferey's (2017) study of adverbial concessive markers in English and French, for instance, the very low score of mutual correspondence (i.e. 15%) between *cependant* and *however* in newspaper articles is partly accounted for by the fact that *cependant* is deleted in translation in 35% of the cases. The analysis of zero correspondences is particularly interesting in that it can shed light on the extent to which different languages need to signal a given discourse relation explicitly by means of a conjunctive marker. Looking at the correspondences of adverbial resultative CMs in English and Swedish, for example, Altenberg (2007) observes that translators translating from Swedish into English delete CMs nearly twice as frequently as translators working in the other direction. Focusing on the Swedish marker *alltså* in translated texts, he then observes that in nearly a third of the cases, this marker has no corresponding equivalents in the English source texts but is added 'out of the blue' in the Swedish translations. Altenberg concludes from these results that there is probably a cultural difference between English and Swedish with respect to the explicit signalling of causal relations: "while causal relationships are often left unmarked in English texts, they tend to be signalled explicitly by a connector to a greater extent in Swedish texts" (ibid.). Other studies which provide valuable information on the differences in cohesive explicitness between languages through the analysis of CM omissions and additions include Aijmer and Altenberg (2002), Fabricius-Hansen (2005), Becher (2011) or Takagaki (2011). Some translation-based studies of CM equivalences also draw attention to cases of 'divergent translations' (Johansson 2007: 25), where a given CM is translated into a unit from a different grammatical class (e.g. *however* translated as French *mais*). These can provide insights into the preferred types of markers used by different languages to signal a relationship between two discourse units (viz. conjunctive adjuncts vs coordinators vs subordinators). In her analysis of the translations of

the Norwegian adverbial additive marker *dessuten* in English and French, Hasselgård (2014a: 74) observes that in English the most frequent translations of *dessuten* are also adverbial markers, whereas in French *dessuten* is most frequently translated by markers belonging to other grammatical classes. This may suggest that French tends to resort to different types of CMs than English and Norwegian to convey discourse relations of addition.

In addition, it is worth noting that some studies have investigated cases of zero correspondences with more theoretical objectives in mind. Recent research has attempted to identify some cross-linguistically valid principles of discourse organisation through the analysis of CM omissions across various European languages. Based on the proportion of zero translations uncovered for CMs expressing a range of logico-semantic meanings, these studies aim to uncover some universal factors determining whether or not a given relation needs to be signalled explicitly by means of a CM (see e.g. Hoek & Zufferey 2015; Zufferey & Gygax 2016; Zufferey 2016; Hoek et al. 2017). For example, using a translation corpus made up of English texts and their translations into four target languages (viz. Dutch, French, German and Spanish), Hoek et al. (2017) observe that conditional relations are the ones that are least often left implicit, irrespective of the target language analysed. They conclude that this relation must be more cognitively complex than additive or causal relations, and therefore more difficult to infer in the absence of an explicit marker (see also Chapter 2 for more on this issue).

Alongside the studies investigating equivalences between CMs by looking at their translations, a number of researchers have also resorted to comparable data – sometimes in combination with translation corpora – to compare the features of conjunctive markers across languages. This type of research aims to explain the patterns of cross-linguistic correspondence – or lack thereof – between CMs by carrying out fine-grained semantic descriptions of these items in original data. Thus, for example, Degand and Pander Maat (2003) analyse three French and three Dutch backward causal CMs (viz. *puisque, car, parce que; aangezien, want, omdat*) in a comparable corpus of news with respect to the degree of speaker involvement (or subjectivity) that they entail. Using samples of 50 occurrences per CM, they rank the six markers on a scale from less to more speaker involvement by identifying the degree of subjectivity of (i) the relation conveyed by the marker and (ii) the segments related by the CM. One of their main findings is that, although French *puisque* and Dutch *aangezien* are generally viewed as equivalent, they differ in terms of the degree of subjectivity that they involve, with *puisque* appearing higher on the speaker involvement scale than *aangezien*. They add that in some cases, the more subjective Dutch marker *want* is therefore a better equivalent to *puisque* than *aangezien*. Similar studies carried out on comparable data include: Pit (2007), who also compares a set of Dutch, French and German

backward causal CMs by reference to the degree of subjectivity that they encode; Lewis (2006), who attempts to account for the lack of equivalence between *on the contrary* and *au contraire* with the help of four parameters, and concludes that the major difference between them is that *on the contrary* always implies the denial of the first segment, whereas *au contraire* does not; Mortier and Degand (2009), who identify the subtle semantic nuances between French *en fait* and Dutch *eigenlijk* by analysing the features of their related segments along the semantic dimensions of factuality, opposition and reformulation, while also complementing this semantic analysis with a study of the translations of the two markers.[10] These studies also frequently include discussions of frequencies of the CMs in original language. The objective is to capture possible differences in the semantic and/or functional scope of the markers compared by contrasting their frequencies in comparable texts. For example, Lewis (2006) observes that *au contraire* is considerably more frequent than *on the contrary* in her corpus (153 vs 29 occurrences per million words), and attributes this difference to the fact that *au contraire* has a wider range of functions than *on the contrary*.

In summary, an extensive body of research has been interested in establishing the degree of equivalence between CMs across languages. The main value of these studies resides in the thorough descriptions that they offer of the markers investigated: they provide illuminating insights into the distinctive features of individual CMs (both within and across languages), making it possible to uncover very fine-grained semantic, pragmatic, stylistic and/or syntactic differences between CMs that may seem equivalent at first sight. Such findings can constitute a valuable resource for language teaching and translator training alike, by pinpointing the factors which should guide the choice of one marker over another, seemingly equivalent one in a given context. By paying attention to cases of zero correspondences, such studies also draw our attention to some (uses of) markers in language A that are best left untranslated in language B.

As explained at the beginning of this section, apart from a few exceptions, the studies described so far usually take the form of case studies, focusing on a limited set of conjunctive markers which they study in great detail. In other words, such research tends to adopt a predominantly semasiological approach to CMs. By contrast, a number of researchers have approached CMs from a more onomasiological perspective: they have attempted to identify language-specific strategies of discourse organisation by including all the CMs expressing a common meaning

---

10.   Lewis's study is based on comparable corpora of political speeches. It is nevertheless included in this overview, as she describes her data as "written-to-be-spoken", i.e. on the border between written and spoken language.

and/or function (e.g. all causal or contrastive markers), instead of starting from a predefined set of forms (see also Crible 2018 on this distinction in relation to the study of CMs).

3.2.2   Onomasiological approaches to conjunctive markers

As compared to the research investigating cross-linguistic equivalences between CMs, onomasiological contrastive work on conjunctive markers is fairly infrequent. One possible explanation for the relative paucity of onomasiological contrastive studies may pertain to the difficulty of establishing a *tertium comparationis* for CMs. As shown in Chapter 2, establishing a reliable TC for the study of conjunctive markers is particularly challenging: the category of markers to be included needs to be circumscribed both functionally and semantically. In addition, as will be discussed in detail in Chapter 6, the delimitation of the segments related by the markers also poses its fair share of difficulties in terms of cross-linguistic comparability. Even in monolingual research, the delimitation of the CM category has aroused considerable controversy. In view of this, it can be very difficult to design a comprehensive list of conjunctive markers expressing a given meaning, which is also comparable across the two or more languages analysed. Questions of cross-linguistic comparability are arguably easier to tackle when they are established on a smaller scale, i.e. for one or two pairs of markers. In such cases, the TC usually relies on (i) a relationship of cognateness between the forms investigated (e.g. *on the contrary/au contraire* in Lewis 2006), (ii) a perceived equivalence between them – either in dictionaries or in the academic literature (e.g. Degand 2004) or (iii) a cursory analysis of translation data, to identify the most common translation equivalents of a given CM – a method that is manageable for a small set of markers, but can become much 'messier' when applied to a larger collection of items (e.g. Mortier & Degand 2009). In sum, issues related to the TC thus probably explain – at least in part – why onomasiological contrastive studies of CMs have tended to remain relatively uncommon.

Studies which adopt an onomasiological approach to the cross-linguistic study of CMs nonetheless exist and can be grouped into two main categories. A first group of studies are mainly methodological in nature: they aim to assess the interoperability of discourse annotation systems and/or taxonomies of discourse relations across languages (see e.g. Zufferey et al. 2012; Iruskieta et al. 2014; Crible & Zufferey 2015; Lapshinova-Koltunski et al. 2015; Zufferey & Degand 2017). The primary objective of these studies is to test whether a given annotation scheme can be applied to a range of languages and, when necessary, to revise the initial taxonomies to improve their cross-linguistic applicability. Such studies typically include markers from a wide variety of semantic categories, analysed in

either comparable or translation data. One excellent example of such research is the study by Zufferey and Degand (2017), who apply the Penn Discourse Tree-Bank annotation scheme, initially designed for the coding of discourse relations in English, to French, German, Dutch and Italian corpus data. Based on the results of a multilingual annotation experiment using the initial coding system, they put forward a revised taxonomy of discourse relations which proves more adapted to the annotation of discourse relations across a range of languages. Although the prime objective of these studies is methodological, they sometimes also shed light on some language-specific discourse features: by underlining cases where a given taxonomy does not apply equally to the different languages under scrutiny, the researchers are able to pinpoint some differences in how these languages tend to express relationships between clauses and sentences.

A second group of onomasiological studies of CMs aim at a broad comparison of the general discourse-organisational strategies of two (or more) languages by analysing one or several semantically-coherent group(s) of conjunctive markers in these languages. The languages are compared along a series of dimensions, including the overall frequencies of explicit CMs that they display, their preferred semantic and/or syntactic types, the positional and combinational patterns of the CMs, etc. The analyses carried out in these studies are very similar to the ones which I intend to perform in this book. Therefore, the remainder of this section focuses more specifically on a few studies of this type, with a view to presenting their main methodological characteristics, as well as the type of results that they can produce.

A first example of a cross-linguistic onomasiological approach to CMs is found in Cuenca (2003), who investigates the use of reformulation markers in a comparable corpus of Catalan, English and Spanish academic papers in linguistics (c. 40,000 words per language). Based on a previously-established list of reformulation markers in the three languages, she extracts all the relevant instances of CMs automatically from the corpus. She observes that reformulation markers are much more frequent in Spanish and, to a slighter extent Catalan, than in English academic texts, in terms of both types and tokens. In addition, she finds that the types of markers used differ across languages: while English tends to prefer simple or grammaticalised reformulation markers (e.g. *or, in other words, that is*), Spanish and Catalan display a preference for more complex or variable markers, such as Spanish *en otras palabras,* which also appears in the following variant forms: *con otras palabras, dicho con otras palabras, formulado in otros términos, con palabras más precisas*. Cuenca attributes the cross-linguistic differences in frequency and use of reformulation markers between English, on the one hand, and Spanish and Catalan, on the other, to diverging rhetorical strategies in the three languages. While English is said to observe more linear and synthetic principles of discourse

organisation, Spanish tends to be wordier, less concise and more digressive in academic writing – which calls for multiple reformulations. Catalan occupies an intermediary position between these two extremes. In a later study, Cuenca and Bach (2007) refine the findings of the 2003 study by investigating the distribution of Catalan, English and Spanish reformulation markers across three semantic subcategories of reformulation (viz. reduction, expansion and permutation). They report that English differs from Catalan and Spanish as regards the subtypes of reformulative relations that it typically signals with conjunctive markers. In addition, they note that, while English tends to have specific markers for each subtype of reformulation, Catalan and Spanish display more polysemous reformulation markers, which can be used to express several subtypes of reformulation. According to Cuenca and Bach (ibid.: 169), this result can again be traced back to the opposition between more vs less concise discourse organisation principles in English as opposed to Spanish and Catalan. Although these two studies rely on fairly small corpora, the methodology that they use is very much in line with the approach that I wish to adopt in this research.

Taboada and Gómez-González's (2012) study of concessive markers in English and Spanish provides another example of a study comparing the frequency and use of a semantically-coherent set of CMs across languages. Adopting the framework of Rhetorical Structure Theory, the authors investigate the expression of concessive relations in English and Spanish across two language varieties, i.e. (written) movie and book reviews and (spoken) telephone conversations. Based on the literature, they compile a list of English and Spanish concessive markers across a range of grammatical categories (e.g. adverbs, coordinators, subordinators, prepositional phrases, and even gerunds such as *admitting that, supposing that*), which they then extract automatically from the corpus. They find that, while written English, spoken English and written Spanish display comparable frequencies of concessive markers, spoken Spanish exhibits a much lower frequency of markers than the other three subcorpora. In terms of CM types, the variety of CMs is much smaller in the spoken than in the written register, in both English and Spanish. In addition to investigating the overall frequencies of concessive markers in English and Spanish, Taboada and Gómez-González's study includes some more 'qualitative' comparisons, i.e. they investigate the preferred order of the related spans across languages and registers and they look at the frequency of combinations of concessive markers (e.g. *but nevertheless*) in English and Spanish. As a general conclusion, the authors state that differences in the expression of concessive relations are in fact more pronounced between registers (or, in this case, modes) than between languages: as they explain, "genre guides and constrains the types of coherence relations used, and […] those constraints are constant across similar genres in different languages" (ibid.: 35). One of the strengths of Taboada

and Gómez-González's study is that they integrate register variation to the cross-linguistic comparison. One of its problems, however, is that the corpora used are once again quite small (i.e. c. 10,000 words per language for the spoken mode; between 62,000 and 90,000 words for the written mode). This calls into question the generalisability of some of their results. For example, for the spoken register, their claims are based on 101 tokens in English, and no more than 24 tokens in Spanish; in addition, many CMs in Taboada and Gómez-González's list are either too infrequent to allow for reliable generalisations to be made, or simply absent from the corpus.

Finally, the studies carried out in the framework of the German-English Contrasts in Cohesion (GECCo) project also constitute an excellent illustration of onomasiological contrastive research into CMs. The GECCo project, carried out by a group of researchers from the University of Saarland (Germany), aims to investigate and compare strategies of cohesion (i.e. not only conjunction, but also reference, ellipsis, substitution and lexical cohesion) in English and German across a range of communicative situations. Like the present book, the GECCo project is grounded in the Systemic Functional framework. With respect to the study of cohesive conjunction, two GECCo studies are of particular interest. In a first article, Kunz and Lapshinova-Koltunski (2014) analyse cohesive conjunction in a written corpus containing original English and German texts from eight different registers, along with their translations into the other language. They aim to investigate the explicit signalling of all logico-semantic relations in English and German. Kunz and Lapshinova-Koltunski report that overall, original German exhibits a higher frequency of CMs than original English, which is in line with the widespread claim that German tends to make discourse relations explicit more frequently than English. These results are corroborated by the analysis of translation data, which reveals a tendency to add CMs in translation from English into German, and to remove them in the other translation direction. The researchers also show that German appears to display a greater variety of CM types than English – a difference which is partly due to the considerable frequency of *and* in the English corpus. With respect to the preferred grammatical types of CMs in each language, the proportion of conjunctive adjuncts (which they call 'conjunctive adverbials') is found to be significantly greater in German than in English (c. 70% vs c. 40%). This difference is often reflected in translation from German into English, where CAs are either left out, or undergo a syntactic shift – typically by being translated by *and*. From a semantic point of view, English and German are shown to both display a preference for signalling additive, causal and temporal relations by means of CMs more frequently than contrastive ones. Nevertheless, German appears to use contrastive CMs more often than English. Finally, English and German are shown to display clear differences in placement patterns of CMs,

especially for CAs, which English places in initial position more than twice as frequently as German. Again, these results are checked against translation data which, interestingly, reveals striking differences in placement between original and translated language for some categories of markers, especially in German. Kunz and Lapshinova-Koltunski conclude that, although English and German have very similar resources at their disposal to signal logico-semantic relationships between clauses and sentences, they display striking differences in the ways in which they exploit these resources.

Information on the differences between English and German uses of cohesive conjunction is also found in Kunz and Lapshinova-Koltunski's (2015) comparison of the use of co-reference, substitution and conjunction in original English and original German across ten written and spoken registers. The results are partly in line with the 2014 study described above, as they point to a more extensive use of cohesive conjunction in German than English, especially with respect to the contrastive relations. In addition, German is once again shown to display a more frequent use of CAs than English, which tends to prefer coordinators. The added value of this study as compared to the one described above is that it adopts a fundamentally variationist approach to the cross-linguistic analysis of cohesion. As opposed to Taboada and Gómez-González (2012), Kunz and Lapshinova-Koltunski demonstrate that contrasts in cohesive strategies are more marked between the two languages than between registers, and that the cross-linguistic differences observed appear to be stable across registers. Thus, at this stage the respective influence exerted by language and register on the frequency and usage patterns of CMs remains a controversial issue. In Kunz and Lapshinova-Koltunski's (2014) study, German is also shown to display more cross-register variation than English, which is comparatively more homogeneous in its use of cohesive devices. Finally, the study reports that language mode (i.e. spoken vs written) plays a significant role on the use of cohesion, with spoken registers exhibiting similar usage patterns, both within and across languages.

The results reported in these studies demonstrate the value of onomasiological studies of CMs, which allow for a multiplicity of very fruitful comparisons. For instance, languages can be compared in terms of the overall frequencies of CMs that they display, with respect to their preferred semantic and syntactic types of CMs, the variety of conjunctive markers used, the positional and combinational patterns of the CMs, and the impact of register variation on all these aspects. From a methodological point of view, these studies tend to be based on comparable data (sometimes complemented with translation-based analyses). This is due to the fact that comparable corpora have been shown to be more reliable than translation corpora for the analysis of CM frequencies and distribution. As optional elements,

CMs tend to be very volatile in translation. More specifically, they have been demonstrated to be especially prone to translation explicitation (see Section 3.1 for a definition), while their use in translated texts has also been shown to reflect the influence of the source language. Accordingly, some researchers have reported significant differences in frequency of CMs across comparable original and translated texts in the same language (see e.g. Cartoni et al. 2011 on English and French; Xiao & Yue 2009 on Chinese). Therefore, relying exclusively on translation data to compare the frequency and distribution of CMs across languages would run the risk of providing a misleading picture of the differences between the languages. This issue will be discussed in greater detail in Chapter 5.

Whereas the English-French language pair was fairly well-represented across the other types of contrastive studies of CMs presented in this section (see e.g. Lewis 2005; 2006; Zufferey & Cartoni 2012 for studies on equivalence; Zufferey & Degand 2013; Crible & Zufferey 2015 for studies assessing the interoperability of taxonomies of discourse relations in English and French), no large-scale corpus-based study has yet undertaken to compare the more general discourse-organisational strategies of English and French written language, in a similar way to Kunz and Lapshinova-Koltunski (2014; 2015) for English and German, or Taboada and Gómez-González (2012) for English and Spanish (but see Crible 2018 for a comparison of English and French CMs in speech). The present study seeks to fill this gap in English-French contrastive research.

3.3   Conjunctive markers in English and French

Corpus-based contrastive research provides little information on the general strategies of discourse organisation of English and French. However, illuminating insights into the differences in frequency and types of markers used in the two languages can be gained from the field of comparative stylistics. This section provides an overview of the research which has specifically compared English and French as regards their frequency of use of CMs (Section 3.3.1), and the types of markers that they tend to choose when signalling a given discourse relation explicitly in text (Section 3.3.2). I show that, although the claims made in these studies remain to be tested on a large empirical basis, to this day they are the only evidence at our disposal regarding the discourse structuration preferences of written English and French – as is clear from the fact that they are still quoted extensively in current English-French contrastive research. As such, they constitute invaluable bases for the formulation of hypotheses concerning the differences between English and French conjunctive marker usage.

3.3.1   Frequency of conjunctive markers in English and French

One of the features of conjunctive markers is that they are said to be linguistically optional, in the sense that the logico-semantic relationships that they signal can in many cases be inferred even in their absence (see Chapter 2). Yet, the extent to which discourse relations need to be signalled explicitly by means of conjunctive markers can vary across languages. Starting from observations on such differences between English and French, a number of researchers have formulated advice for translators on how to handle cohesive devices when translating either from French into English, or from English into French. The dominant claim in these studies has been that French tends to be more explicitly cohesive than English. Hervey and Higgins (1992: 49), for example, explain that "[i]t is more common in French than in English for texts to be explicitly structured with connectors ('or', 'donc', 'ainsi', 'en effet', 'par ailleurs', 'en revanche', etc.) that signpost the logical relationships between sentences". Likewise, according to Armstrong (2005: 196) "the general tendency seems to be to mark cohesion in French in a more explicit way than in English, using more linguistic material". In the same vein, Vinay and Darbelnet (1995: 234) associate each language with a different mode of textual development: while English is said to typically adopt an intuitive development, French rather follows a rational development. This opposition between English and French textual strategies is outlined as follows (ibid.):[11]

> An intuitive development tends to leave maximum autonomy to each segment of the message. This autonomy corresponds to a situation in which the sequences between events are not always visibly linked by a relationship of causality. On the other side, a rational development tries to place the relationships which unite the segments of the utterance into a logical sequence […]. French, at least in its literary, philosophical, scientific and legal discourse, cultivates these structural markers and finds it difficult to manage without the connections they can bring to the presentation of thought. English, on the other hand, even in its traditional written forms, is much less dependent on these explicit connectors and instead relies on the juxtaposition, or parataxis, of the elements of the utterance, leaving it to the readers to provide for themselves the necessary connections.

Similar statements on the frequency of use of explicit CMs in English and French are legion in the literature: they are also found in Van Hoof (1989: 39), Grellet (1991: 177), Martel (1993), Poncharal (2005; 2007), Delisle (2013: 603, 643) or Quillard (1997: 321) – who goes so far as to state that French texts "swarm with logical connections" as compared to English texts [my translation]. In view of

---

11.   Vinay and Darbelnet's *Stylistique comparée du français et de l'anglais* was first published in French in 1958. I have used the English translation of this book, which was published in 1995.

these differences, many of these authors instruct translators to comply with the preferences of the target language in terms of discourse organisation, by inserting CMs when translating from English into French – as in Example (1), where *en effet* is added in the target text – and removing markers when working in the other translation direction – as in (2), where both *donc* and *or* are removed in the English translation.

(1)   EN. But one feels that this is an abnormal condition which lacks the elements of healthy growth, the growth that augurs eventual stability. Socially and politically, there is [ø] widespread discontent […].

FR. Cette situation toutefois ne laisse pas de sembler anormale par l'absence apparente des éléments essentiels d'un progrès sain, avant-coureur d'un équilibre définitif. Il règne **en effet** dans les masses […] un profond mécontentement, un malaise politique etc. […].

(Vinay & Darbelnet 1995: 239)

(2)   FR. Pour que la proposition française soit retenue, il faudrait **donc** qu'elle soit acceptée par les 28 autres négociateurs. **Or**, aux dernières informations, elle aurait, au mieux, l'appui de 5 ou 6 délégations.

EN. In order to be adopted, the French proposal would [ø] have to be accepted by the other 28 negotiators. [ø] At the last count it could rely on the support of five or six delegations at most.

(Chuquet & Paillard 2017: 41)

Poncharal (2007: 118), for example, explains that when translators faithfully reproduce the cohesive structure of English source texts into French, the target texts usually seem incoherent, and become difficult to read. Conversely, Hervey and Higgins (1992: 49) insist that "an English TT [target text] using explicit connectors to reproduce all those found in a French ST [source text] is likely to seem tediously over-marked in discourse structure, and therefore stilted, pedantic or patronizing". In order to ensure intersentential links, English is said to rely on different means, including exact repetitions of lexical words across sentences (i.e. lexical cohesion), or the use of anaphoric pronouns (i.e. cohesive reference; see e.g. Quillard 1997; Poncharal 2005; 2007; Chuquet & Paillard 2017: 40).

While the assertion that French tends to make a more extensive use of explicit CMs than English is widespread in the contrastive literature, some researchers have actually made the opposite claim, i.e. they have stated that English generally signals logico-semantic relations explicitly by means of CMs, whereas French displays a tendency to simply juxtapose discourse segments without specifying the relation between them. Such a view is held by Guillemin-Flescher (1981: 83, 189), who notes that explicit relational markers are frequently added in translation from French into English, and removed in the other translation direction ("en anglais,

il faudra tout d'abord expliciter la relation" (ibid.: 189)). Likewise, Chuquet and Paillard (1987: 148–150) mention a tendency of English to use explicit markers of relations between clauses, where French rather tends to juxtapose them. This phenomenon, which is also noted by Gallagher (1995) and Mason (1998), is exemplified in (3) to (5).

(3)   EN. Tom went to the door **and** began to softly lift the latch; **then** he pressed gently, **and** the door yielded a crack; he continued pushing cautiously, **and** quaking every time it creaked.

FR. Tom s'approcha de la porte ; avec précaution il souleva le loquet, poussa légèrement ; la porte grinça ; il continua à pousser prudemment, tremblant d'angoisse à chaque fois que la porte grinçait.

(Guillemin-Flescher 1981: 83)

(4)   FR. Il était temps : les coups de téléphone des provinces, optimistes de minuit à deux heures, commençaient à ne plus l'être.

EN. It was high time, **for** the reports telephoned in from the provinces […] were beginning to sound a different note.

(Chuquet & Paillard 1987: 150)

(5)   FR. Elle aurait passionnément souhaité faire avec lui un pèlerinage aux Metz, où ils avaient été si heureux ; il choisit d'y aller sans elle, en octobre 1837, pour s'y trouver seul avec leurs souvenirs.

EN. She would have loved passionately to make with him a pilgrimage to Metz, where they had been so happy, **but** he chose to go there without her, in October 1837, in order to be alone with their shared memories.

(Gallagher 1995: 214)

Thus, the research currently available provides us with contradictory claims on the frequency of use of CMs in English and French. This is most probably because the studies mentioned above suffer from a series of methodological weaknesses. A first major problem pertains to their relative lack of empirical foundation. Several of these studies largely rely on introspection to formulate statements on the differences between English and French, resorting to limited sets of isolated, hand-picked examples to illustrate their claims (e.g. Hervey & Higgins 1992; Van Hoof 1994; Vinay & Darbelnet 1995; Poncharal 2005). A case in point concerns the French CMs *or, en effet* and *ainsi*: based on the observation that these items are rarely translated into English, some researchers have made the sweeping generalisation that French tends to use more explicit CMs than English. Yet, as demonstrated by Zufferey (2016: 276), intrinsic properties of individual markers may cause them to undergo frequent omissions in translation, which does not necessarily mean that this tendency can be extended to the language as a whole. Thus, while the claims made in these studies are certainly informative, the extent to which they can be

generalised is questionable. Moreover, although these researchers admittedly had an excellent command of both English and French, compounded by their experience in the practice of translation and/or translation teaching for this language pair, intuition is infamous for its unreliability, especially when it comes to frequencies of linguistic phenomena (see Section 3.1.1).

In addition, some of the studies discussed above, while not resorting strictly to intuition, rely on fairly small amounts of data to make very general assertions about the languages as wholes. Guillemin-Flescher (1981), for example, bases most of her conclusions on the French novel *Madame Bovary* and two of its translations into English, although she sometimes complements this 'corpus' with "some examples from contemporary writers […] and […] some non-literary utterances" (ibid.: ix, my translation; see also e.g. Chuquet & Paillard 1987; Martel 1993; Gallagher 1995 for similar research designs). These studies usually provide little or no information on the methodology used, and do not offer any quantifications of the linguistic phenomena investigated. In this sense, while the researchers usually refer to their data set as a 'corpus', these studies may not be viewed as corpus-based in a strict sense, i.e. as defined in Section 3.1.1. Although they arguably allow for more reliable claims than the studies relying chiefly on intuition, it seems as though much larger amounts of texts, accompanied by transparent frequency quantifications of the linguistic items under scrutiny in each language, are required to permit generalisable conclusions on the differences between languages. In addition, it should be noted that most of these studies rely exclusively on translation data – often unidirectional. However, as briefly underlined in Section 3.2, translation data has to be approached with some critical distance when it comes to the comparison of CM frequencies (see also Mason 2001: 72; Chapter 5 for a detailed discussion). Finally, the data analysed in these studies is generally restricted to a single text type. Yet, as demonstrated by Mason (2001) and Leroux (2012), the propensity of one language to signal a given discourse relation explicitly, as well as the differences between languages in that respect, may vary according to the communicative situation. Leroux (2012), for example, observes that differences between English and French in the explicit signalling of causal relations are not stable across the three text types included in her study, viz. newspaper articles, blog posts and nationalist political speeches. She concludes that "in contrastive linguistics, it is difficult to talk about English and French globally" (ibid.: 19). Instead, researchers should minimally be aware that a difference observed in one situation of communication will not necessarily hold in different contexts. All in all, contrastive studies of CM frequencies in English and French thus broadly belong to the first generation of contrastive studies described in Section 3.1.1: unfortunately, they have not yet integrated the numerous possibilities afforded by a corpus-based approach to contrastive analysis.

A second important limitation of the studies available is that they usually make very general claims on the differences in frequency of English and French CMs, without considering the possibility that these differences may be sensitive to the type of logico-semantic relation expressed by the markers. Yet, cross-linguistic differences in degree of cohesive explicitness have been shown to be partly dependent on the type of relation investigated. In her contrastive study of textual organisation in French and Japanese, for example, Takagaki (2011: 247) explains that, while the hypothesis that French uses more explicit CMs than Japanese is confirmed for the causal and temporal relations, with respect to relations of contrast and addition, French appears to resort to juxtaposition more frequently than Japanese. In a comparable corpus of academic articles, Cuenca and Bach (2007) uncover a larger frequency of reformulation markers in Spanish than in English, which stands in opposition with previous contrastive findings on English and Spanish discourse organisation strategies. They argue that the direction of the difference uncovered in their study is probably specific to the relation of reformulation, and may be traced back to broader cultural differences between Spanish and English, viz. a tendency towards digression and redundancy in Spanish, and towards conciseness and synthesis in English (see Section 3.2.2 for more details). In a similar vein, Moreno (1998: 550) warns against the potential danger of putting all the relations expressed by conjunctive markers "in the same bag" when comparing discourse structuring strategies across languages. She argues that "broad generalizations like the fact that more or less connectives, or text markers, are used in a language in comparison with another […], though interesting at an exploratory level, are of little help in fields such as teaching or translation" (ibid.). These studies therefore all underline a necessity of qualifying the contrastive claims on the use of CMs according to the type of logico-semantic relation. Also note that in a number of studies, it is not entirely clear what types of conjunctive devices are covered by the contrastive claims made by the authors: does the larger frequency of explicit conjunctive markers in French as compared to English pertain to all types of discourse-structuring devices (viz. coordinators, subordinators and conjunctive adjuncts), or just one or two of these categories?

With respect to the relation of contrast, a few studies have specifically zoomed in on the behaviour of contrastive conjunctive markers in English-French translation. However, these studies once again offer contradictory insights. Gallagher (1995) and Mason (1998), for example, assert that French usually juxtaposes discourse segments standing in a contrastive relationship, whereas English requires that relations of contrast be signalled explicitly by means of conjunctive markers. They explain that, in translation from French into English, it is often necessary to restore the explicit links that were absent from the original: as Mason (1998: 172) notes, "[i]n translation into English […], straightforward transfer of

the ellipsis [of the CM] may create problems of coherence for target-text readers and consequently […], there is a tendency for translators to restore explicit junction". Martel (1993), on the other hand, claims the opposite. She observes that French contrastive CMs occur twice as frequently as their English equivalents in original fictional texts. Likewise, while Guillemin-Flescher (1981: 189) notes a tendency of French additive markers to be translated by contrastive markers in English, Ballard (1995: 260–261), who investigates the other translation direction, discusses some instances of English *and* which are translated by French markers of contrast. Such counter-examples are also provided by Mason (2001: 72–73). Again, these discrepancies are most likely due to the fact that these studies tend to rely on very small data sets to formulate general contrasts on the use of CMs in English and French (e.g. Martel's study is based on about ten pages of Camus's *L'Étranger*, and one chapter of Hemingway's *To Have and Have not*).

In summary, the evidence currently available on the differences in frequency of CMs in English and French remains largely inconclusive and riddled with contradictions. Faced with the many indeterminacies characterising the field of contrastive stylistics, Mounin (1976: 232–233, quoted in Ballard 1995: 226–227) states that no definite answers on cross-linguistic contrasts will be obtained "until the issue is tackled using the only method that can do away with subjective intuitions and overall impressions: a statistical method […]. [W]e must do what other domains have started to do, i.e. choose a corpus and count occurrences" [my translation]. As already stressed earlier, with respect to the possible differences in degree of cohesive explicitness between English and French, such an approach is still lacking: no study has yet undertaken to systematically compare English and French on the basis of large, written corpus data. One of the key objectives of this book is to fill this gap.

Despite the various weaknesses underlined in the previous pages, it must be highlighted that early English-French contrastive studies remain highly valuable for present-day contrastive research. While the claims that they contain are neither based on large corpora, nor accompanied by any rigorous quantifications of the phenomena observed, they result from years of practice in the fields of translation and/or translator training, during which their authors gained first-hand experience of the problems facing translators when trying to transpose content from one language to the other in the most idiomatic way possible. In view of the relative lack of methodological elaboration of these studies, the cross-linguistic differences that they identify are not directly generalisable. However, as stated by Butler (2004: 150), whereas "introspection is a very unreliable way of getting at details of how people really behave linguistically", it constitutes an essential starting point for the formulation of hypotheses on language use, which can then be tested against corpus data. Accordingly, the studies discussed in this section constitute excellent

bases for the formulation of hypotheses to be tested on a larger scale. More generally, these studies are unique with respect to the breadth of the descriptions which they offer of the contrasts between English and French: far from simply focusing on the differing strategies of discourse organisation of English and French, they address an impressive diversity of aspects of the two linguistic systems (e.g. the lexicon, tense, aspect, metaphorical meaning, word order, prosody, etc.). To this day, they remain the most wide-ranging overviews of the differences between English and French – which undoubtedly explains why they remain very influential today.

Finally, for the sake of completeness, it ought to be specified that some translation-corpus-based studies focusing on other features of CMs in English and French have sometimes, in passing, provided some clues on the possible differences in cohesive explicitness between the two languages. In a study aiming to identify a range of factors influencing the translation equivalences of concessive CMs across languages, Dupont and Zufferey (2017: 286) note that omissions of CMs are nearly twice as frequent in the French-English as in the English-French translation direction. We suggest that this may provide evidence in favour of the claim that French tends to use more explicit CMs than English. In their methodological study aiming to test the homogeneity of original and translated texts in terms of frequency of causal CMs, Cartoni et al. (2011) report four times as many CM additions in French as a target language as in the English target texts, which is also consistent with the hypothesis of a more extensive use of CMs in French than in English. Zufferey and Cartoni (2012: 247), by contrast, in their study of equivalences between causal markers in English and French, observe that, whereas translators working in the French-English direction tend to maintain CMs, translators working in the other direction use more varied equivalents of English CMs, including paraphrases and zero correspondences. Even though they are only mentioned in passing, and do not provide sufficient evidence to give a definite answer to the questions raised here, these results are highly relevant for the present study. Finally, in a very recent study, Crible (2018) compared the use of discourse markers (i.e. a category consisting of conjunctive markers as defined here, along with a number of speech-specific discourse particles such as English *well, you know*; French *ben, tu vois, hein*, etc.) in a comparable corpus of speech across eight registers, using a fully onomasiological approach. Crible reports a significantly higher frequency of discourse markers in French than in English. Her large-scale corpus-based study can thus be said to provide corroborating evidence in favour of the statements found in the intuition-based contrastive literature – although it is important to stress that not all the markers included in her study perform a cohesive function (cf. e.g. a range of discourse particles such as *oh, sort of, hein,*

bah). Despite the contribution of Crible's study, there is still considerable scope for an equivalent study on written language, especially as the literature provides ample evidence that the frequency and use of conjunctive markers vary extensively across language modes (see e.g. Altenberg 1984; Altenberg 1986; Taboada & Gómez-González 2012; Kunz & Lapshinova-Koltunski 2015).

3.3.2   Preferred types of conjunctive markers in English and French

As stated by James (1980: 113), "while every language has at its disposal a set of devices for maintaining textual cohesion, different languages have preferences for certain of these devices and neglect certain others" (see also Chapter 2 on this). In addition to the claims on the differences in frequencies of CMs in English and French, the studies discussed in Section 3.3.1 make a series of claims on the types of structural devices that each language tends to prefer in order to convey explicit links between two events or states of affairs. In this respect, the statements found in contrastive linguistics are far more consistent than those concerning the frequency of explicit CMs in the two languages. There is fairly widespread agreement in the literature that, when they are in fact made explicit, relations between clauses tend to be signalled by means of coordination in English, and subordination in French. Chuquet and Paillard (1987: 151), for example, state the following: "[w]hen relationships between processes are signalled explicitly in both languages, they are frequently realised through coordination in English, and subordination in French" [my translation]. Along the same lines, Delisle (2013: 601) asserts that "English has a predilection for juxtaposition and coordination, whereas French naturally prefers articulation and subordination" [my translation]. Similar statements – all of which are translation-based – are found in Guillemin-Flescher (1981: 143), Newmark (1988: 59), Ballard (1995), Quillard (1997: 322) or Hoarau (1997: 196). Accordingly, these authors provide numerous instances of English coordinators translated by French subordinators – as in (6) – and vice-versa – as in (7). It is worth noting that a great majority of these shifts in syntactic ordering (or 'agencement syntaxique', to use Guillemin-Flescher's terminology) reflect diverging preferences of the two languages, rather than being triggered by systemic constraints. In other words, translating these syntactic structures literally will rarely result in ungrammatical utterances. However, according to Delisle (2013: 130), these language-specific preferences are strong enough that translating interclausal links directly from one language to the other can lead to cumbersome, unidiomatic and poorly organised texts. He uses the phrase 'anglicismes syntaxiques' [syntactic anglicisms] (ibid.: 602) to talk about literal transfers of the textual structuration of English to French target texts.

(6)   EN. Some thieves work in pairs or groups. In these instances, one or more distract the victim by some means **and** another accomplice does the actual stealing.

FR. Les voleurs travaillent parfois à deux ou en groupe. **Pendant que** l'un attire l'attention de la victime par quelque moyen, un autre la vole.

(Delisle 2013: 603)

(7)   FR. Il arriva à la gare **pour** s'apercevoir qu'il avait oublié son portefeuille.

EN. He got to the station **and** realised he had forgotten is wallet.

(Chuquet & Paillard 1987: 152)

It is also important to specify that the claims described here pertain to all types of coordination and subordination, and not only those that are signalled by an explicit marker. For example, one type of French subordination that is said to frequently correspond to coordination in English is participle clauses, as in (8).

(8)   EN. Camel was sitting on his desk **and** surveying the scene with obvious enjoyment.

FR. **Assis** sur la table de travail, Camel observait ce spectacle avec un plaisir manifeste.   (Ballard 1995: 262)

Once again, the statements outlined above need to be approached with caution, as they are also affected by the methodological shortcomings characterising the comparisons of CM frequencies discussed in Section 3.3.1. However, unlike the claims on the degree of explicitness of the two languages, the introspection-based assertions on English and French preferred types of conjunctive markers have been revisited on an empirical basis. The research carried out by Cosme (2006; 2008a; 2008b), in particular, has relied on authentic comparable and translation corpora to test the claims provided in the intuition-based contrastive literature. The results of her corpus analyses only partly substantiate these claims. Based on a comparable corpus of newspaper editorials, for example, Cosme (2008a; 2008b) shows that, while French does appear to be markedly less fond of coordination than English, no significant difference is found between the two languages in their propensity to resort to subordinating constructions. Translation-corpus findings, on the other hand, provide firmer confirmation of the statements found in contrastive stylistics: Cosme observes both a majority of shifts from coordination to subordination in translation from English into French, and a majority of subordination to coordination shifts in the opposite translation direction. She confirms that an overwhelming majority (viz. 97.4%) of these shifts are optional and stem from preferences of the respective languages, instead of systemic differences between them (Cosme 2008a: 96; Cosme 2008b: 357). In both the comparable and the translation data, Cosme also reports some variation in the functional types

of clauses that give rise to differences in strategies of clause linking. For example, among the most frequent types of French subordinating clauses corresponding to English coordination, we find relative clauses, participle clauses, and adverbial finite clauses – most frequently temporal, as in Example (9).

(9)   FR. Il avait disparu pour se dissoudre dans la lumière jaune, pâle et indécise **tandis que** la porte se refermait.

EN. He was gone, dissolving in the pale, uncertain, yellow light, **and** the door was closing.   (Cosme 2008a: 105)

Zooming in on the interclausal coordinators *and* and *et*, Cosme (2006) reports that, whereas interclausal *and* is significantly more frequent than its French equivalent in a comparable corpus of fictional texts, no significant difference in frequency is uncovered in a comparable corpus of newspaper articles. These results thus, once again, only partially confirm the claims found in the non-corpus-based contrastive literature and suggest that the difference between English and French may be register-dependent. Turning to the analysis of translation shifts involving *and* and *et*, on the other hand, Cosme uncovers a higher frequency of shifts from coordination to subordination in the English-French than in the French-English translation direction, which again confirms the claims found in the literature.[12]

In summary, the results presented in Cosme's work tend to confirm the claims formulated in the non-corpus-based contrastive literature, although they show that these need to be qualified to some extent: the greater tendency of French to use subordination is not confirmed by the analysis of comparable data, and the patterns of correspondence between English coordination and French subordination appear to be sensitive to (i) the type of relation holding between the related clauses and (ii) the communicative situation. By basing her cross-linguistic comparisons on a solid empirical foundation, Cosme can be said to have provided a very reliable picture of the preferred patterns of clause linking in English and French. It may therefore seem superfluous to raise this question again in this book. However, Cosme looks at all types of coordinating and subordinating structures, both functionally (e.g. adverbial, relative, participle clauses) and semantically (e.g. additive, contrastive, temporal interclausal relations). It is therefore not clear how French and English fare concerning the expression of logico-semantic relations

---

12.   From a methodological point of view, it is interesting that, in all three studies, the results obtained by Cosme are not fully identical in the comparable and the translation data. This reaffirms the necessity of being well-aware of both the potential and the limitations of each type of corpus when comparing languages, as the cross-linguistic differences emerging from translation and comparable corpora are not always fully consistent (see Section 3.1.2 on the advantages and shortcomings of comparable and translation data).

of contrast between clauses. In this respect, for example, Chuquet and Paillard (2017: 164) suggest a reverse pattern of correspondence between English and French for the specific subordinator (*al)though*, which is frequently translated by coordinators or conjunctive adjuncts in French, as in (10). After analysing a sample of about 5,000 occurrences of English and French coordinators extracted from fictional texts, Hoarau (1997: 96) states that the pair *but/mais* is the one which poses the fewest translation problems, although it is not entirely clear what she means by this (do they cause fewer shifts? Or are the shifts simply more regular?). Further research is therefore required in order to shed light on the English and French preferred patterns of clause combination for meaning relations of contrast.

(10)   EN. **Though** three stable cats were kept to deal with outside rats she was the only house cat.

FR. Trois autres chats avaient pour mission de pourchasser les rats dans les étables, **mais** elle seule était autorisée à pénétrer dans la maison.

(Chuquet & Paillard 2017: 164)

Finally, with respect to the expression of discourse relations by means of conjunctive adjuncts (e.g. *however, cependant*), some researchers have claimed that French tends to show a greater preference for such discourse-structuring devices than English. For example, Vinay and Darbelnet (1995: 237–242) put forward a categorisation made up of four formally-defined categories of CMs – labelled from A to D. Category A comprises conjunctive adjuncts such as *aussi* or *cependant*, along with other types of discourse-organising cues such as *j'en viens maintenant à* or *comme nous l'avons déjà vu.* Category B includes referential expressions such as *this, those, ceux, ceci*. Category C includes what Vinay and Darblenet refer to as 'simple linking connectors', and includes coordinators (e.g. *et, ou, ni, mais, donc*), but also some conjunctive adjuncts such as *d'abord, ensuite*. Finally, category D includes cases of zero markers, where the related clauses are simply juxtaposed. Vinay and Darblenet (ibid.: 238) claim that CMs included in category A are more frequent in French than in English. Note, however, that the delimitations of their categories are strikingly vague (in particular regarding the boundaries between categories A and C, see also Ballard 1995 for a criticism of Vinay and Darbelnet's categorisation), and it is therefore not perfectly clear how to interpret Vinay and Darbelnet's statement. A perhaps clearer assertion is provided by Chuquet and Paillard (2017: 40), who state that "French tends to use argumentative connectors such as *en effet, or, en revanche, par ailleurs…* where English simply uses *and* or *but*, or even juxtaposes the two sentences" [my translation]. More generally, among the authors who have insisted on the greater tendency of French than English to resort to explicit markers of discourse relations, quite a few focus more particularly, or sometimes even exclusively, on expressions of these relations

through CAs (see e.g. Van Hoof 1989; Hervey & Higgins 1992), which suggests that this type of CM is particularly frequent in French as compared to English. Similarly to most of the claims discussed so far, however, these statements largely remain to be verified empirically.

In summary, with respect to the types of explicit conjunctive markers that tend to be preferred by each language, the contrastive literature suggests that English has a greater tendency to resort to coordinators to express logico-semantic relations between discourse segments, while French more commonly uses subordinators and conjunctive adjuncts for that purpose.

3.4   Conclusion

The aim of this chapter was to provide a broad overview of the corpus-based contrastive research on conjunctive markers. The chapter started with a general presentation of the field of corpus-based contrastive linguistics, discussing the main benefits that can be gained from a corpus approach to the comparison of languages. The two major types of corpora used in contrastive linguistics (viz. comparable corpora and translation corpora) were presented, along with the main advantages and limitations associated with each type of data. Finally, I presented a recent strand in corpus-based contrastive linguistics, which has underlined the importance of taking register variation into account when comparing languages.

The second part of the chapter gave a state-of-the-art survey of corpus-based contrastive research on conjunctive markers. Two main groups of studies were identified. Firstly, a number of studies have been concerned with establishing the degree of cross-linguistic equivalence between conjunctive markers. Starting from the observation that few one-to-one equivalences may be established between CMs across languages, these studies have tried to identify the factors – whether semantic, stylistic or syntactic – making it possible to predict the best equivalent(s) for a given CM in a specific context. Accordingly, such studies typically provide very detailed descriptions of the markers investigated and of the fine-grained differences between CMs, both within and across languages. They are typically – although not exclusively – based on translation data, and most frequently take the form of case studies, focusing on a limited set of markers described in great detail. A second group of (less frequent) studies adopt a more onomasiological approach to CMs, analysing the use of coherent groups of markers expressing a given meaning and/or function, instead of focusing on a well-defined set of forms. Within this category, we find (i) methodological studies attempting to test and/ or improve the cross-linguistic validity of taxonomies of discourse relations by applying a given annotation scheme to a range of languages; and (ii) studies which

aim at a broad comparison of the strategies used by different languages to signal a specific (set of) discourse relation(s) in text. As the present study corresponds to the second category of onomasiological approaches to CMs, particular focus was laid on such studies, with a view to presenting their major methodological features, as well as the main type of results that they have produced. It was shown that such research can give rise to very fruitful comparisons of the differences in preferred discourse-organisational strategies of two or more languages, by analysing phenomena as varied as the number and type of explicit CMs typically used in different languages, the positional and combinational patterns of the CMs, the impact of register on CM use, etc. No study of this type has yet been carried out for English and French written language.

The final part of this chapter discussed a number of contrastive studies on English and French which, although they are not based on corpus data, provide relevant insights into the use of conjunctive markers in the two languages. These studies have made claims on (i) the differences in degree of cohesive explicitness between English and French and (ii) the preferred types of markers used by each language to signal the relationships between two clauses. With respect to the frequency of CMs, most studies have stated that French tends to be more explicitly cohesive than English, although some researchers have also made the opposite claim. In terms of the preferred types of markers used by the two languages, most researchers seem to agree that French tends to resort to subordination and conjunctive ajuncts, while English prefers coordination to relate two clauses or sentences. Due to a series of methodological weaknesses (e.g. a lack of empirical foundation, or a disregard for the fact that differences in CM use may be sensitive to the type of logico-semantic relation expressed by the markers), the evidence provided in these studies remains somewhat inconclusive, and their claims still have to be verified empirically. Such research nevertheless provides a valuable basis for the formulation of hypotheses to be tested on a larger scale. This is what the present study intends to do, for the specific logico-semantic relation of contrast.

CHAPTER 4

Systemic Functional Linguistics, corpus linguistics and the textual metafunction

This chapter deals with the relationship between the frameworks of Systemic Functional Linguistics and corpus linguistics, which are both central to this book. It is made up of two main parts. Section 4.1 focuses on the benefits that can be gained from the combination of the theoretical framework of Systemic Functional Linguistics with the methodology of corpus linguistics. Section 4.2 then zooms in on one particular aspect of the SFL theory which the present study intends to further develop through the application of corpus methods, viz. the textual metafunction, and in particular the system of thematic structure, approached from the perspective of conjunctive marker placement.

4.1 Systemic Functional Linguistics and corpus linguistics: A promising synergy

This study adopts a combined Systemic Functional and corpus approach to the study of conjunctive markers of contrast. This section sets out to make clear the advantages of such a disciplinary synergy, which has recently been gaining ground in contemporary linguistic research. Section 4.1.1 starts by discussing some areas of common ground between SFL and corpus linguistics, which have contributed to creating "a natural affinity between systemic theory and corpus linguistics" (Halliday 2006: 293). Section 4.1.2 turns to differences between the two approaches, and shows that SFL and corpus linguistics can be viewed as complementary. Finally, Section 4.1.3 provides a brief overview of the progress which has been made towards the rapprochement of Systemic Functional Linguistics and corpus linguistics over the years.

4.1.1 Systemic Functional Linguistics and corpus linguistics: A 'natural affinity'

The combination of the Systemic Functional theoretical framework with corpus linguistics (henceforth: CL) has been advocated by eminent researchers in SFL,

such as Butler (2004) or Thompson and Hunston (2006a). One major argument provided in favour of such an approach has been that the two frameworks share central features and concerns that make their juxtaposition particularly promising and even 'natural', to use Halliday's (2006: 293) words.

A first striking point of convergence between SFL and CL is their insistence on the importance of basing linguistic descriptions on naturally-occurring, authentic data (see e.g. Thompson & Hunston 2006b: 4; Hunston 2013: 618; Teich 2009: 113). As made clear in Chapter 3, one of the four defining features of a corpus – at least in the strict sense of the term, as defined in corpus linguistics – is that it must contain authentic texts. Likewise, the reliance on naturally-occurring data has played a central part in SFL from the very beginning (Matthiessen 2006: 103). But whereas in corpus linguistics, occurrences of linguistic phenomena in authentic texts are typically regarded as valuable in their own right, in order to understand the position of textual data in the Systemic Functional framework, it is useful to provide a brief description of the concept of instantiation. SFL is primarily concerned with the description of the system of language, defined as its overall underlying potential as a meaning-making resource (Halliday & Matthiessen 2004: 26). Halliday and Matthiessen (ibid.: 27) describe the system as an abstract, or even 'virtual', theoretical representation of the possibilities offered by a given language to create meaning. One implication of the view of the language system as an abstract entity is that it cannot be observed directly. Instead, in Systemic Functional Linguistics, the system is said to be "'instantiated' in the form of text" (ibid.: 26). In other words, texts are seen as concrete 'realisations' of the potential contained in the abstract system of language. The relationship between system and text is such that "language as system can only be observed through text" (Bednarek 2010: 241). This means that the only way to provide a theoretical account of the system of language is to observe naturally-occurring textual data instantiating that abstract system, hence the central position of authentic data in the SFL theory.[1]

A second aspect of common ground between SFL and CL pertains to the importance that they attach to corpus frequencies for language theorisation. As discussed in Chapter 3, one of the many advantages of corpus data is that it allows for quantifications of the observed linguistic phenomena, leading to reliable generalisations

---

1. Halliday has frequently used the analogy of meteorology to illustrate the relationship between system and instance (e.g. Halliday 1991; Halliday 1992; Halliday & Matthiessen 2004). He explains that system and text are related to each other in the same way as the climate and the weather: "texts are the day-to-day weather patterns displaying variations in temperature, humidity, air pressure, wind direction, and so on, all of which can be observed, recorded and measured" (Halliday 1992: 66). Based on all these weather patterns, the climatologist is able to draw general conclusions about the climate of a given area.

on the frequencies of use of forms and functions in specific contexts. Such poten-tialities of corpora are in perfect accordance with the view of language put forward by SFL, where the linguistic system is essentially conceived of as a system of choices. More specifically, the linguistic system is made up of a set of subsystems, which offers language users with a multiplicity of mutually exclusive choices – generally leading, in turn, to more fine-grained choices – every time they express themselves (Bloor & Bloor 2004: 3). For example, whenever a writer or speaker wants to relate two clauses in order to form a clause complex, they will need to choose whether to do so by means of parataxis or hypotaxis. Simultaneously, they will have to pick the type of logico-semantic relationship that they want to express between the clauses. At the same time, the speaker/writer will also need to choose what tenses and voices (passive vs active) to use for the verbs in each clause, select the polarity of the two clauses (positive or negative), and so on. In SFL, the choices offered by the language system are represented in the form of system networks, as in Figure 7, which illus-trates the systems of choices associated with clause complexing.
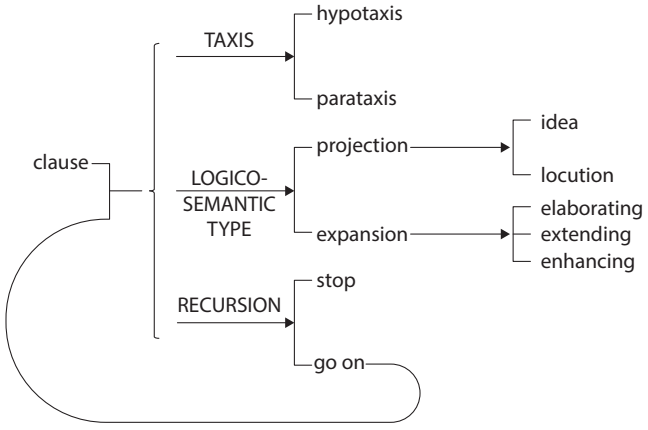
Figure 7. The systems of clause complexing (Halliday & Matthiessen 2004: 373)

The link between language viewed as a system of choices, on the one hand, and corpus frequencies, on the other, is that very early on, SFL has insisted on the importance of determining the probabilities associated with the different choices within a system, highlighting that these choices are far from always equiprob-able. In some systems – which Halliday (1992: 65) calls 'skew' – one choice will appear much more probable (or 'unmarked') than its alternative(s). Halliday and James (1993), for example, establish that the probability of choosing a positive over a negative polarity for a given clause is of about 9 to 1. Thus, from the start, SFL has put forward a view of language as an "inherently probabilistic system"

(Halliday 1991: 31), where statements on linguistic choices in terms of 'either/ or' need to be balanced in terms of 'more/less likely' choices (Nesbitt & Plum 1988: 8). For Systemic Functional linguists, the probabilities associated with linguistic choices form an integral part of the description of a language: as stated by Nesbitt and Plum (1988: 8), "[p]atterns of typical choice are as much part of the grammar of a language as are the choices themselves" (see also Halliday 1991: 31). Against this backdrop, it becomes clear that corpora and corpus frequencies have an important role to play in the description of language: as argued by Nesbitt and Plum (1988: 17), "probabilization of the grammar can only proceed by being empirically based on the sampling of text". In other words, in the same way as texts are viewed in SFL as instantiations of the system, corpus frequencies are perceived as instantiations of the probabilities associated with each choice in the grammar of the language (Halliday 1991: 31; Tucker 2006: 94; see also Halliday 2005 for a collection of Halliday's most relevant work on the role of quantification in SFL).

A number of early SFL studies have used authentic data to define the probabilities associated with the choices available in various linguistic systems (see e.g. Halliday & James 1993 on the systems of primary tense and polarity; Plum & Cowling 1987 on tense choice). With respect to the topic of this book, the study by Nesbitt and Plum (1988) is of great relevance: based on a collection of 123 interviews containing as many as 2,733 clause complexes, they establish that clauses are more likely to be related through: (i) parataxis (70%) than hypotaxis (30%), and (ii) expansion (84%) than projection (16%) (see Chapter 2 for definitions of each component of the system of clause complexing; see also Figure 7 for a visual representation of this system). In addition, Nesbitt and Plum show that, although the systems of taxis (i.e. parataxis vs hypotaxis) and logico-semantic relations are *grammatically* independent of each other (in the sense that the choice of a given option in one system does not affect the possible choice of options in the other system), they are *statistically* dependent: depending on the choice made in one system, the various options available in the other system will not appear equally likely. For example, within the subsystem of expansion, the probability of clauses being related paratactically instead of hypotactically is of about 3 to 1 for relations of elaboration, whereas this ratio is reversed for relations of enhancement. According to Nesbitt and Plum, these results provide evidence in favour of an interaction between the systems involved in clause complexing (see also Matthiessen 2002; 2006; Caffarel 2006 for similar, later studies).[2] All in all, although Nesbitt and Plum

---

2.   Incidentally, this type of analysis is very much in line with the kind of quantitative analyses advocated in present-day corpus linguistics, where linguists are encouraged to apprehend linguistic phenomena as being multifactorial, and consider the possible interactions between linguistic features (see Chapter 8 for more on this).

do not explicitly claim to situate themselves in the field of CL, some of the methods that they use are strikingly similar to those used by CL.

A third area of commonality between SFL and CL pertains to their concern for the influence of the social context on language use: in both SFL and CL, register variation is given a key role in linguistic description (see e.g. Hunston 2013: 619; McDonald & Woodward-Kron 2016: 160). In CL, the study of register variation was mainly initiated by Biber (see e.g. Biber 1988; 1995), whose approach is predominantly inductive (Neumann 2014: 37; Teich 2003: 27): starting from a set of linguistic features analysed in texts, he uses statistical methods to determine their patterns of co-occurrence, which results in so-called 'dimensions' of variation. These clusters of co-occurring linguistic features – or dimensions – are then interpreted in terms of the communicative functions to which they correspond, such as 'involved vs informational' or 'narrative vs non-narrative' (Biber & Conrad 2001: 184). Thus, for example, the combination of high frequencies of first- and second-person pronouns, questions, contractions, final prepositions and present-tense verbs will be associated with 'involved' linguistic productions, whereas large frequencies of nouns, prepositions, longer words, along with a high type-token ratio correspond to a more informational focus (ibid.: 185). For each dimension, registers are ranked according to how they behave with respect to the collection of linguistic features associated with that dimension.

Register is also a central component of Systemic Functional theory. In fact, Halliday was one of the first researchers to delve into the influence of context on language use. Compared to Biber, who investigates registers "only from the text end" (Teich 2003: 27), SFL adopts a more deductive, or top-down approach to register analysis (Neumann 2014: 37). Register is defined as a variety of language as determined by its situational context. The communicative situation is therefore the main point of departure here and is itself modelled as a configuration of features along three dimensions of variation, viz. field, tenor and mode (Thompson 2014: 40). The field of discourse broadly refers to the subject-matter of the linguistic activity, i.e. what is being talked about; the tenor has to do with the social characteristics of the interactants and the relationships between them, and thus belongs to the interpersonal plane of language; finally, the mode of discourse is essentially concerned with "how the language is functioning in the interaction" (ibid.), and includes parameters such as the channel of communication – i.e. speech vs writing; the degree of interactivity of the text – viz. monologic vs dialogic; and the rhetorical function of the text – e.g. persuasive, expository, didactic (see Halliday & Hasan 1989; Neumann 2013 for detailed overviews). The configuration of contextual features along these three dimensions will prompt language users to make particular choices from among the set of possibilities offered by the language system. In this sense, just as in Biber's text-based approach, register in SFL is conceived of as a frequency-based phenomenon: depending on the

context in which language is being used, the probability of users making one or the other choice will vary, sometimes substantially. With respect to the mood system of grammar, for example, the probability for the writer to choose the imperative mood over the declarative one will be very different depending on whether we are looking at recipes, news reports, academic writing, etc. (Thompson 2014: 41). Therefore, registers can be defined as subsystems that display their own sets of probabilities with respect to the range of choices offered by the linguistic system: as Halliday (1992: 68) puts it, "register variation is, simply, variation in the setting of grammatical probabilities". Once again, the only way to gain access to these register-specific probabilities is through the examination of the frequency of linguistic features in authentic texts representing a range of registers.

In summary, SFL and CL share a number of central tenets, among which the importance attached to the analysis of authentic language, the conviction that corpus frequencies are crucial for language theorisation, or the focus laid on register variation. In fact, Halliday (2006: 293) goes as far as to say that SFL and CL "share most assumptions and aims, and notions of scholarly practice, differing just in the resources they bring with them, or in their strategies or priorities in the use of such resources". These areas of common ground have led him to consider the two frameworks as standing in a "symbiotic and synergistic relationship" (ibid.). Sharoff (2017: 534) even speaks of a "genetic link" between SFL and CL. As explained earlier, the close relationship between the two approaches has been manifest from the start in the SFL literature, which contains many allusions to the value of using corpus methods in view of theorising the language system. One excellent example is Halliday's (1991) article entitled 'Corpus studies and probabilistic grammar', which he concludes by stating that "corpus studies have a central place in theoretical investigations of language" (ibid.: 41). Despite the apparent affinity between the two approaches, however, the effective combination of SFL and CL has been relatively slow to prosper. One possible explanation for this might be that, in addition to their similarities, SFL and CL also display notable differences in the ways in which they approach language.

4.1.2   SFL and CL: Areas of divergence

In addition to the similarities described in the previous section, SFL and CL also display some significant differences in the ways in which they approach linguistic description, which have sometimes made the proponents of one approach suspicious of the other. As made clear by Thompson and Hunston (2006b: 1–4), the major part of the tensions between SFL and CL essentially boils down to the distinction between theory, on the one hand, and method, on the other. As is probably clear from what precedes, SFL takes a predominantly theory-oriented perspective

on language. The ultimate goal of Systemic Functional linguists is to provide an all-encompassing, consistent account of the workings of the linguistic system. While SFL has always insisted on the importance of resorting to authentic data, the use of such resources has tended to remain somewhat limited, and generally comes second with respect to theoretical description, instead of being given full priority, as is generally the case in corpus linguistics (see Section 4.1.3 for more details). This situation is summarised by Thompson and Hunston (2006b: 2) as follows:

> Although the theoretical framework has been developed based on naturally-occurring examples […], until relatively recently the amount of data analysed has been comparatively small, and corpora have been used to test, rather than to drive, the theoretical description.

Faced with the apparent 'tidiness' of Systemic Functional theoretical descriptions, corpus linguists sometimes have the impression that "linguistic analyses have to be 'shoehorned' into the existing categories" (Thompson & Hunston 2006b: 3). For them, this can come at the expense of the accuracy of linguistic descriptions, which may consequently fail to do full justice to the patterns actually observed in the corpus data (Hunston 2013: 618). In other words, as stated by Thompson and Hunston (2006b: 3), for the adepts of the corpus approach, the feeling can be that "aspects of language that cannot be accounted for in terms of the three-part system [of metafunctions] are not accounted for at all" (see also Butler 2004: 171). This is partly due to the fact that everything in the SFL system is so tightly interconnected, that making one slight change to the system in order to account for the patterns observed in the corpus can have the effect of disrupting the balance of the entire structure (Thompson & Hunston 2006b: 3). It is true that the 'messiness' of corpus data – which contrasts with the aforementioned tidiness of theoretical descriptions – can prove very unsettling for Systemic Functional linguists, who find themselves compelled to rethink, sometimes fundamentally, their traditional view of language (Butler 2004: 168).

As opposed to SFL, corpus linguistics is first and foremost a methodology.[3] Accordingly, one main criticism that Systemic Functional researchers have addressed to corpus linguists is the lack of theoretical elaboration characterising

---

3. Note that opinions sometimes diverge as to whether corpus linguistics should be viewed 'simply' as a methodology, or whether it can be conceived of as a linguistic discipline (or even a theory of language) in its own right. Many corpus linguists nevertheless agree that, in spite of the many theoretical *implications* that it may have (both upstream, with the design of corpus studies being determined by theoretical factors; and downstream, with the results of corpus analyses usually leading to a partially new conception of some aspects of the linguistic system), corpus linguistics essentially remains a methodology (see e.g. McEnery et al. 2006: 7–8; Taylor 2008; Gries 2009: 1 for a discussion of this issue).

their work. For Thompson and Hunston (2006b: 2), for instance, "[c]orpus lin-guistics […] is almost perversely theory-thin, and leads to 'thin' descriptions". In a more measured way, Halliday and Matthiessen (2004: 35) deplore the fact that, by laying too much emphasis on the methodological aspects of their work, corpus linguists have sometimes "encourage[d] [other researchers] to believe that there is some disjunction between data-gathering and theorizing". The separation between data and theory operated by CL is epitomised by the corpus-driven methodology, which rests on the idea that researchers should try to emancipate themselves as much as possible from prior preconceptions or theories on language, in order to fully prioritise what the data has to say (see Tognini-Bonelli 2001). Such a concep-tion of linguistic analysis clearly conflicts with the SFL perspective on language, where corpus instances are viewed as manifestations of a higher-order theoretical system. For Systemic Functional linguists, one major problem stemming from the relative lack of theoretical grounding of corpus work is that the field broadly con-sists in a large collection of detailed descriptions of individual 'bits and pieces' of a language, which are very interesting in themselves, but usually difficult to relate with one another in order to obtain a more general, coherent picture or theory of language. This issue is summarised by Thompson and Hunston (2006b: 3) as fol-lows (see also Halliday & Matthiessen 2004: 35; Butler 2004: 157):

> CL […] has typically been more like Anouilh's '*voyageur sans bagages*': the relative absence of theoretical elaboration has meant that new insights and changes in one area are easier to incorporate in the description, since they have little impact in the other areas. This can generate a satisfying sense that rapid advances are being made. On the other hand, it is sometimes difficult to see where those ad-vances are leading […]. Detailed descriptions of individual 'bits of language' are fascinating, but it is often difficult to relate them to an overall picture of language, let alone a social semiotic theory. In gross terms, then, we have a version of the conflict between 'woods' and 'trees'.

Importantly, it must be underlined that the view of corpus linguistics as 'theory-thin' presented here is the one taken in some of the SFL literature, as a way of high-lighting the disparities between SFL and CL. In reality, the picture is much more complex than this. Especially in recent years, many corpus linguists have sought to articulate the corpus methodology with pre-existing theories on language (see e.g. the recent rapprochement between corpus-based research and Construction Gram-mar or Second Language Acquisition). In fact, it would be more accurate to state that, as a methodology, CL is not tied to any specific theory of language, and can therefore be applied to virtually any of them (see e.g. McEnery et al. 2006: 7 on this).

The differences between the SFL and the CL approaches to linguistic descrip-tion may appear quite substantial. However, instead of being daunted by the dispari-ties between them, a number of scholars have recently argued for a complementarity

between the theory-oriented perspective of SFL and the methodological approach of CL (see among others: the edited book by Thompson & Hunston 2006a; Butler 2004; Bednarek 2010; Hunston 2013). These researchers have stressed that the differences between SFL and CL are such that the weaknesses of one approach can be complemented by the strengths of the other: SFL could provide the theoretical grounding that is not always present in corpus linguistics, whereas corpus linguistics might contribute the solid, stylistically-varied empirical foundation that is so much in line with some of the basic precepts of SFL, together with powerful methods of extraction and analysis. What is advocated by these researchers is a dialectical relationship between SFL and CL, in which "each perspective interpenetrates with and constantly redefines the other" (Halliday & Matthiessen 2004: 36). More precisely, as Tucker (2006) explains, the corpus data must serve to inform and refine the theory, so that it can provide a full account of the functions and meanings observed in the data; on the other hand, the theory in turn needs to give shape to the apparent 'messiness' emerging from the corpus data (see also Teich 2003: 29; Neumann 2013: 4). As argued by Butler (2004: 163,175), the challenge resides in finding the right balance between the SFL theory and the CL methodology so that, while corpus results should not be swept under the carpet just because they call into question the well-established SFL theoretical framework, the SFL theory must also not be thrown away altogether to "build up descriptions of languages piece by piece" in a wholly corpus-driven manner (ibid.: 163). Instead, as Butler (ibid.: 176) states:

> [W]e need an ambitious research programme in which corpora are used to tackle head-on the task of investigating which aspects of [the theory], including the most fundamental, we genuinely need to abandon in the face of the new evidence, which of them can remain, but with perhaps quite extensive modifications, and which are still robust enough to withstand the onslaught with only minor changes.

This view could be said to represent some sort of middle ground between (i) an approach where the corpus is merely seen as "a repository of examples to back pre-existing theories or a probabilistic extension to an already well defined system" (Tognini-Bonelli 2001: 84), and (ii) a fully corpus-driven approach where, as mentioned earlier, "the commitment of the linguist is to the integrity of the data as a whole" (ibid.), which needs to be approached "with a minimum of interference from preconceived ideas about the lexicogrammar" (Butler 2004: 154).

Naturally, such a synergy as the one recommended here is made all the easier by the numerous points of convergence between SFL and CL described in Section 4.1.1. These recent developments are also fully in line with Halliday's initial hopes for SFL: with his Systemic Functional framework, Halliday wished to break the mould of traditional grammar which, he thought, "was a

subject with too much theory and too little data" (Halliday 1992: 61). From the beginning, he insisted on the importance of relying on as much data as possible to describe the linguistic system: as he explained, the larger the corpus, the more accurate the account of the system was going to be (ibid.: 67). Halliday (2005: 173) viewed linguistics as being no different from any other scientific field, where the accumulation of new evidence leads to more accurate theoretical descriptions:

> [A]fter all, that's what it did in physics, where more data and better measuring transformed the whole conception of knowledge and understanding. How much the more might we expect this to be the case in linguistics, since knowing and understanding are themselves processes of meaning.

For the sake of completeness, it should be noted that the opposition between theory and methodology is not the only aspect along which SFL and CL diverge. Other differences between the frameworks include: (i) a tendency to prioritise grammar in SFL, and lexis in CL – although both frameworks highlight that lexis and grammar are not to be viewed as separate phenomena (Hunston 2013: 622); or (ii) a primarily paradigmatic and hierarchical modelling of language in SFL – embodied by the view of language as a system of choices ('either/or'), with choices in one system giving access to more fine-grained choices in other systems – as opposed to a syntagmatic and linear view of language use in corpus linguistics – where one of the main concerns pertains to recurring patterns of co-occurrence of linguistic units (Tucker 2006: 91). As they are less relevant with respect to the present study, these differences are not developed further in this chapter.

4.1.3   Corpus-based Systemic Functional Linguistics: Where do we stand?

Sections 4.1.1 and 4.1.2 provided a number of arguments in favour of the combination of SFL and CL, showing that the two approaches are both similar and complementary. Despite the fairly strong case for a synergy, and the very early insistence, in the SFL literature, on the interest of adopting a corpus methodology for grammatical descriptions, the combination of SFL and CL has nevertheless been fairly slow to catch on. The situation is summarised by Teich et al. (2006: 247) as follows:

> SFL has a strong tradition in text analysis and interpretation. Its focus is on register variation, where a register is said to be characterized by the greater-than-random co-occurrence of particular linguistic features. However, with few exceptions […] SFL has so far primarily worked with text samples instead of corpora, thus missing the chance of further developing its theory in the direction of probabilistic grammar.

In other words, for a long time, most researchers in Systemic Functional Linguistics have kept basing their linguistic descriptions on the manual analysis of fairly limited bodies of authentic texts, sometimes referred to as 'collections' or 'archives' (Butler 2004: 165; Teich 2003: 27; Bednarek 2010: 238). As explained by Butler (2004: 151), collections or archives of texts are fundamentally different from corpora in the strict sense, i.e. as defined in Chapter 3, and in corpus linguistics in general: in addition to being comparatively small, and analysed fully manually, they are "not necessarily ordered, the selection of material need not be made on linguistic criteria, and there may be no claim that the material is in any sense representative of a language, or one of its varieties". In addition, in SFL, authentic textual data has frequently been used either as a mere repository from which good examples were extracted to illustrate the theoretical claims being made (as in Halliday & Matthiessen 2004; Lavid et al. 2011; Banks 2017) or, more commonly, to provide a detailed description of a particular area of language through the lens of the SFL theory (see Butler 2004: 169). In the latter case, the theory, whose general validity is usually taken for granted, is most often applied to the data, rather than being truly (re)shaped by it in the dialectical fashion advocated above (Butler 2004: 169; Thompson & Hunston 2006b: 2). All in all, for a long time, most SFL research has failed to exploit corpus resources to the full of their potential.

Butler (2004: 166–168) attempts to identify some of the reasons that might explain the somewhat limited recourse to corpora in SFL. Alongside the researchers' fear of seeing their certainties over the organisation of the system overturned in the face of corpus data – a factor that was already mentioned in the previous section – the relative lack of corpus work in SFL can be traced back to the paucity of tools allowing for automatic analyses of grammar. In corpus linguistics, lexis is usually the point of departure of the analysis. This makes the automatic retrieval of the relevant linguistic material fairly easy since, as explained in Chapter 3, the corpus methodology is particularly well-suited for form-based types of research. In SFL, on the other hand, the focus is on the description of grammar (Hunston 2013: 622). Although SFL and CL both view grammar and lexis as a single phenomenon spread along a continuum (see e.g. Halliday 1992: 63 on this), the starting point of SFL linguistic analyses is generally grammatical phenomena, which are much harder to retrieve automatically from corpus data (Sharoff 2017: 537). It is, for example, very difficult to automatise the analysis of information structure, i.e. the division of information between Given and New throughout the linear flow of discourse. This partly explains why Systemic Functional linguists have tended to limit themselves to small samples of texts analysed fully manually (e.g. by reading the whole texts and manually identifying the items analysed, instead of at least extracting the linguistic phenomena under scrutiny automatically; see

Butler 2004 for a detailed overview of the factors accounting for the lack of corpus research in SFL).[4]

Nevertheless, in recent years, notable efforts have been made towards greater collaboration between SFL and CL. One decisive initiative in this respect was the book entitled *System and Corpus: Exploring Connections*, edited by Thompson and Hunston (2006a), which contains thirteen studies attempting to bridge the gap between the two approaches.[5] Since then, an increasing number of studies have worked in the direction of a tighter integration between SFL theory and corpus methods, with some Systemic Functional linguists getting further acquainted with the ins and outs of the corpus methodology, and some corpus linguists looking at their data through the prism of SFL concepts and theories.[6] This set of studies spans a wide range of linguistic phenomena, including register analysis (e.g. Neumann 2013; 2014; Teich 2013; Miller & Johnson 2013), thematic structure (e.g. Hasselgård 2004a; Lavid 2010; Herriman & Bostrom Aronsson 2009; Berry et al. 2014), cohesion and clause complexing (e.g. Kunz & Steiner 2012; He & Yang 2015; He & Wen 2017), interpersonal and evaluative meanings (e.g. Bednarek 2010; Miller & Johnson 2014), phraseology (e.g. Holtz 2007; Tucker 2009; Hunston 2013) or discourse analysis (e.g. McDonald & Woodward-Kron 2016; Zinn & McDonald 2018). All these studies meet the requirements of the corpus approach, in that they base their SFL analyses on large bodies of authentic data which they – most often – analyse automatically. It nevertheless ought to be noted that they differ in the extent to which they implement the dialectical process between theory and data described in the previous section, with some studies still using corpus data predominantly to apply the SFL theory rather than to truly challenge some of its categories. In parallel to these studies, we have seen the emergence of a number of corpus tools specifically tailored for Systemic Functional analyses, such as the SysCon, SysTag and SysFan office suite, which enables annotations at various

---

4.   This is of course not to say that corpus work does not imply some amount of manual analysis. In fact, most corpus studies (including studies on discourse, such as this one) involve an extensive, usually pain-staking process of manual work (see e.g. Chapter 6 on issues of disambiguation and coding). However, one defining feature of corpus studies (in the strict sense of the term) is that they should involve the use of computers to automatise at least part of the analytical process (if only for extracting the potentially relevant items from the corpus, without which the analysis of large amounts of data would simply not be feasible).

5.   The book was edited as a result of the 29th International Systemic Functional Congress, held at the University of Liverpool in July 2002.

6.   This is not to say that such studies did not exist before; but Thompson and Hunston's book gave a noticeable boost to this strand of research, by explicitly underlining the advantages, as well as tackling some areas of difficulty that such a combined approach entails.

levels of the linguistic system (see Wu 2009); or the UAM Corpus Tool (O'Donnell 2009), which allows for a multilayer annotation of the data in the form of system networks (see also Teich et al. 2006; Teich 2009; Wu 2009; Sharoff 2017 on the main corpus resources available for SFL). However, there is still room for improvement in this respect, notably in terms of the degree of automation that these tools permit (see e.g. Teich et al. 2006; Teich 2009 on this).

Interestingly, a notable number of studies have used a combined SFL and corpus approach to compare languages. Examples include Teich's (2003) comparison of original and translated texts in English and German across a wide array of SFL categories (e.g. information and thematic structure, grammatical metaphor, mood, transitivity, etc.); Neumann's (2013; 2014) work on register variation in English and German original and translated data; Hasselgård's (2004a) and Lavid's (2010) comparisons of thematic choices in English and Norwegian, and English and Spanish, respectively. With respect to cohesion in general, and conjunctive markers in particular, the studies carried out in the framework of the GECCo project are exemplary of the type of approach adopted in the present study. The project aims to compare the use of various cohesive strategies in English and German, using both original and translated data across a range of written and spoken registers (see e.g. Kunz & Steiner 2012; Kunz & Steiner 2013; Kunz & Lapshinova-Koltunski 2014; Kunz & Lapshinova-Koltunski 2015; Kunz et al. 2017; see also Chapter 3 for a more detailed discussion of the GECCo project). In all these contrastive studies, the combined SFL and CL approach is shown to be ideal for the comparison of languages, in that (i) the Systemic Functional framework allows for detailed descriptions and comparisons of the possibilities (or choices) offered by the linguistic systems under scrutiny; and (ii) the quantitative corpus approach makes it possible to pinpoint differential uses (or instantiations) made of linguistic systems that may seem similar at first sight (Kunz & Steiner 2012: 225–226; see also Teich 2003: chap. 3 for a detailed discussion of the advantages of SFL for contrastive analysis). These potentialities of the combined SFL, CL and contrastive approach are fully in line with the objectives of this study, which aims to provide a maximally detailed picture of the differences between English and French uses of CMs of contrast, both in terms of systemic possibilities and frequencies of use.

The present book situates itself in the set of recent studies seeking to both benefit from *and* contribute to the combination of Systemic Functional Linguistics and corpus linguistics. Although it is well on track, the synergy between the two approaches is still underway. On the one hand, a number of SFL studies claiming to adopt corpus methods still fail to take full advantage of what CL has to offer: for example, they keep relying on fairly small data sets (e.g. analysing a single text, see Goatly 2004; Hunston 2013) and/or performing fully manual analyses of their data (which may, once again, be attributed to the lack of resources for the SFL analysis

of corpus data; see e.g. Matthiessen 2006). Likewise, not all CL studies referring to SFL are yet fully grounded in the Hallidayan theory: instead, some studies have initiated the transition by 'borrowing' some core SFL concepts – such as thematic structure or cohesion – either to define their object of study, or to interpret their corpus results (see e.g. Hasselgård 2010b: 11, who explains that her analysis "borrows some terms and definitions from […] Halliday"; Altenberg 2006, who draws upon the SFL systems of information and thematic structure to interpret his corpus results on CA placement in English and Swedish; see also Flowerdew & Forest's 2015 corpus study of signalling nouns for a similar approach). Thus, all the studies that combine SFL theory and corpus methods are far from equal in the extent to which they actually integrate the two approaches, and very often, one framework clearly dominates over the other. One of the aims of the present study is to take a further step towards greater collaboration between the two frameworks.

4.2   Zooming in on the textual metafunction

The objective of this section is to set the scene for the analyses to be carried out in Chapter 8, which aims to investigate and compare the placement patterns of conjunctive adjuncts in English and French. Section 4.2.1 discusses a number of theoretical concepts in SFL which are central for the analyses of conjunctive marker placement. It first defines the textual metafunction, and then focuses more particularly on one of its subparts, viz. the system of thematic structure, which revolves around the notions of Theme and Rheme. Section 4.2.2 then explores the links between the systems of thematic structure and conjunction.

4.2.1   The textual metafunction and thematic structure: Theme and Rheme in Systemic Functional Linguistics

One of the central tenets of Systemic Functional Linguistics is its view of language as an inherently social system (e.g. Halliday 1978) that is used to perform three major functions – called metafunctions of language: (i) to refer to entities in the outside world – such as objects, people, circumstances, events – or in our own consciousness – e.g. ideas, feelings; (ii) to enact social relationships between the participants in the communication; and (iii) to organise messages in a way "that makes them fit smoothly into the unfolding language event" (Thompson 2014: 145). These metafunctions are labelled (i) ideational; (ii) interpersonal; and (iii) textual, respectively, and each of them is associated with its own set of systems of choices in the grammar. For example, the ideational metafunction offers a number of choices with respect to the type of process expressed in a given clause (e.g. mental – *think, believe*; verbal – *say, claim*; material – *run, shower*, etc.), which

in turn determine choices pertaining to the number and types of participants in that process (e.g. a verbal process will necessarily involve a 'sayer', and may also entail the explicit mention of a so-called 'receiver' – the one to whom the saying is addressed – and a 'verbiage' – i.e. what is being said. Mental processes, by contrast, will involve a 'senser' – i.e. the person who experiences the mental process; and a 'phenomenon' – what is in fact experienced; see e.g. Thompson 2014: 94–111; Bloor & Bloor 2004: 106–126 for overviews of the types of processes and participants). The interpersonal metafunction, for its part, determines choices such as the mood (viz. declarative, interrogative, imperative) and the polarity (viz. positive vs negative) of the clause. Finally, the textual metafunction consists of three main systems, viz. (i) cohesion – discussed at length in Chapter 2; (ii) information structure – i.e. the division of the message into Given and New information; and (iii) thematic structure – i.e. the linear organisation of the information in the clause across the two functional units of Theme and Rheme.

In Systemic Functional Linguistics, each clause is seen to realise the three metafunctions simultaneously: for each clause, language users need to select a process together with relevant participants in that process; they also need to decide whether to formulate that clause in the declarative, imperative or interrogative form; in addition, they have to make choices concerning the information that they are going to present as recoverable from the surrounding context (i.e. Given) and the information that should be viewed as newsworthy (or New) by the receiver, etc. As a result, any clause can be analysed with respect to each of the three metafunctional strands: in fact, Systemic Functional linguists suggest that in order to fully account for the meanings expressed by a given clause, it is necessary to analyse it through the lens of each of the metafunctions that it realises (Eggins 2004: 210–213). The focus of this book is on the textual metafunction, which has been described by SFL researchers as the enabling function of language: it is the metafunction that actualises the ideational and interpersonal meanings by organising them in a way that is as effective as possible given the purposes and the context of the linguistic activity (Eggins 1994: 273). In this section, I concentrate on the textual system of thematic structure: this system is the most directly relevant with respect to the placement of conjunctive markers, since it is concerned with the way constituents are ordered in the clause (Eggins 1994: 273). Thematic structure is also the system that holds centre stage in most SFL accounts of the textual metafunction as, according to Halliday and Matthiessen (2004: 105), "thematic organization of the clauses […] is the most significant factor in the development of the text".

4.2.1.1   *Theme and Rheme in Systemic Functional Linguistics*
In Systemic Functional Linguistics, thematic structure is the linguistic resource enabling the clause to function as a message – i.e. as an organised unit which fits

into the flow of discourse (Halliday & Matthiessen 2004: 64). Thematic structure is made up of two major constituents, called Theme and Rheme. The Theme is defined functionally as "the element which serves as the point of departure of the message [and] which locates and orients the clause within its context" (ibid.: 66).[7] In one of his early works, Halliday (1970: 161) also described the Theme metaphorically as "the peg on which the message is hung". As such, in both English and French, the Theme corresponds to the clause-initial position: it can be formally identified as the first clause element with a function in transitivity (viz. the first subject, object, complement *or* adjunct in the clause), together with all the elements that precede it (e.g. Hasselgård 2004b: 65; Banks 2017: 45; see below for further discussion of the delimitation of the Theme). It ought to be noted, however, that the Theme is not *defined* by its position, but simply *realised* by it. The Theme is a fully functional notion, and in some languages, its realisation can take very different forms (e.g. in Japanese, the particle *-wa* is used to assign thematic status to whatever precedes it; see Halliday & Matthiessen 2004: 64). The second component of thematic structure, viz. the Rheme, corresponds to everything that follows the Theme. It is the part of the message in which the Theme is developed. Thus, in Example (1), for instance, the Theme corresponds to the subject of the clause (*Kuala Lumpur*), while all that follows this subject is analysed as the Rheme.

(1) Kuala Lumpur	is the capital of Malaysia. (Bloor & Bloor 2004: 73)
THEME	RHEME

It is worth noting that, in its most typical realisations, thematic structure tends to map onto information structure, in the sense that the information presented in the Theme typically corresponds to the Given information – i.e. the information that is presented as already known by the receiver – whereas the rhematic part of the clause generally hosts the information that is presented as newsworthy, and which is in accordance with the communicative goals towards which the text is oriented (see e.g. Halliday & Matthiessen 2004: 93; Fries 1994). As Halliday (1967: 205) explains, however, this pattern of congruence is "merely the unmarked sequence", and there may be a number of reasons why writers or speakers may wish to present the New information at the beginning of the clause (see e.g. Halliday & Matthiessen 2004: 93–94 for a discussion of some of the rhetorical effects that can be achieved

---

7.   Note that in earlier accounts, Halliday suggested that the Theme could also be defined as 'what the clause is about' (e.g. Halliday 1967; Halliday 1985). This view has been criticised quite vividly, however (see e.g. Downing 1991; see also Gómez González 2001: 115–126 for a discussion), and it is now fairly well-accepted that Theme and topic should be viewed as two different units.

by disrupting the unmarked association between Theme/Given and Rheme/New). For this reason, and as opposed to previous approaches to thematic and information structure (e.g. the Prague School's Functional Sentence Perspective, see Fries 1995 for a comparison with SFL), SFL treats Theme/Rheme and Given/New as two clearly independent – although closely interrelated – systems of choices. It is also noteworthy that the Theme and the Rheme have been shown to differ in their typical 'grammar', that is, in the type of linguistic forms and meanings that they generally contain. For example, Themes have been demonstrated to typically host referential information, realised in the form of nominal groups. Rhemes, by contrast, most frequently convey evaluative information. Formally, they may be composed of elements as diverse as nominal, verbal, adverbial or prepositional groups. The Theme and the Rheme also differ in their preferred ways of ensuring textual cohesion: whereas Themes tend to contain a high frequency of conjunctive markers, Rhemes typically realise cohesion by means of lexical ties. Cohesive reference, by contrast, is frequent in both subparts of the message (see e.g. Fries 1994; Hasselgård 2004b; Cummings 2009 for more details). In the remainder of this section, I provide more detailed descriptions of these two notions of Theme and Rheme, which are central to describe the ways in which constituents are distributed across the clause.

4.2.1.2   *Several types of Theme*

The part of thematic structure that has by far received the most attention in the SFL literature is the Theme: due to its important discursive status – as the point of departure of the message – the Theme has been studied extensively, and several subtypes have been identified. Firstly, although in Example (1) above, the Theme consists of a single element (corresponding to the subject of the clause), it is very frequent for Themes to include several components, performing different functions in the construction of the message. As Halliday and Matthiessen (2004: 79) make clear, a Theme *must* always contain one – and *only* one – experiential element, i.e. one element that refers to entities in the outside world (such as participants or circumstances, e.g. *you, the house, last week*), and plays a function in the transitivity of the clause (cf. above).[8] This experiential element is referred to as the 'topical Theme'. It marks the end of the Theme, and everything that comes after it is considered to be part of the Rheme (Eggins 1994: 277). However, the topical Theme may be preceded by one or several elements which have no function in transitivity, and as such do not exhaust the 'thematic potential' of the clause.

---

8.   Except for a number of special themeless structures, see Halliday and Matthiessen (2004: 98–100)

These elements may be of two types: (i) textual Themes are elements which signal explicitly how the clause relates to the preceding discourse, and commonly correspond to conjunctive markers (e.g. *and, but, however, furthermore*); (ii) interpersonal Themes typically provide information about the stance of the writer or speaker towards his/her message (e.g. *honestly, sadly, clearly*), or on the nature of the interaction between the participants (e.g. vocatives or forms of address, such as *John, Sir, darling*). When a Theme contains textual and/or interpersonal elements preceding the obligatory topical Theme, it is referred to as 'multiple Theme'. The typical – although by no means obligatory – order of Themes in a multiple Theme is the following: textual ^ interpersonal ^ topical (where ^ means 'followed by'; see e.g. Rossette 2009: 11). Examples of multiple Themes in English and French are provided in (2) and (3). Interestingly, as the reader may have noticed, the complete structure of the Theme mirrors the three metafunctions of language (viz. textual, interpersonal and ideational) described earlier in this section[9] (Eggins 1994: 271).

(2)	But	surely	the course	doesn't start till next week. (Thompson 2014: 164)
	Textual	Interpersonal	Topical	Rheme
	Theme			Rheme

(3)	Mais	probablement	toutes les filles	ont des idées pas ordinaires (Caffarel 2006: 171).
	But	*probably*	*all girls*	*have funny ideas*
	Textual	Interpersonal	Topical	Rheme
	Theme			Rheme

The second major distinction with respect to the Theme is that between marked and unmarked Themes. For each kind of mood structure in the language system (viz. declarative, imperative, interrogative), one type of thematic choice – or starting point – appears as the most typical among a range of options. In declarative clauses, for example, the most typical point of departure is the subject, as in Example (1). In *wh*-interrogative clauses, on the other hand, the most typical starting point for the clause tends to be the *wh*-word, as in (4) (Thompson 2014: 150). When the topical Theme of a given clause conflates with the most typical choice for that clause type, it is said to be unmarked. By contrast, when other experiential elements occupy thematic position, the Theme is said to be marked. In declarative clauses, for example, the most typical example of a marked Theme are fronted adjuncts, as in (5). Complements may also occur as the Theme of declarative

---

9.   The topical function is a subpart of the ideational metafunction.

clauses (as in (6)), although they appear a lot more marked than thematic adjuncts (see e.g. Butt et al. 2000: 140–141; Thompson 2014: 150 on the cline of markedness with respect to thematic choices). In addition, the language system also offers a number of marked thematic structures which make it possible to grant particular prominence to a given constituent chosen as Theme. These include so-called 'thematic equatives' – otherwise known as pseudo-cleft sentences, see Example (7) – or 'predicated themes' – commonly referred to as cleft sentences outside SFL, as in Example (8) (see e.g. Thompson 2014: 153–159 for a summary of the main 'special thematic structures' available in English). The purpose of this section is not to provide a full account of all the types of thematic structures identified in SFL. Whenever relevant for the analysis of the corpus data, however, other, more specific cases of thematic analysis will be discussed in Chapter 8.

(4)	Who	wants a glass of white wine? (Halliday & Matthiessen 2004: 76).
	THEME	RHEME

(5)	For a long time,	the Spartans proved themselves invincible on land. (Bloor & Bloor 2004: 76)
	MARKED THEME	RHEME

(6)	Particularly significant	was the way the subjects reacted to the third task (Thompson 2014: 149).
	MARKED THEME	RHEME

(7)	What really annoyed me	was that they didn't tell me the truth (Thompson 2014: 154).
	MARKED THEME	RHEME

(8)	It was Diana	who had donated blood 36 times (Eggins 1994: 295).
	MARKED THEME	RHEME

According to Thompson (2014: 148), unmarked Themes correspond to the constituents chosen as Theme "unless there are good reasons for choosing something else". In other words, as highlighted by Halliday (1967: 213), the selection of a given Theme for a clause is a meaningful choice: the decision to grant thematic status to one element instead of another testifies to the method of development chosen for the text. It generally reflects a wish by the writer (or speaker) to draw attention to a particular element in the clause, or create a specific effect – such as highlighting a contrast between two textual units, or emphasising a shift to a new stage in discourse (see e.g. Caffarel 2006; Martin & Rose 2007: 192). The type of contrastive effect that can be achieved by the choice of Theme is quite clear in Example (9) below, where the thematisation of the two locative adjuncts

underlines the contrast between what happens *in Switzerland* and what happens *here*. Interestingly, because different thematic choices are not equivalent in terms of the discourse effects that they accomplish, choices pertaining to thematic structure are likely to vary across registers (see e.g. Caffarel 2006: 183–192; Thompson 2014: 177–180).

(9)	[I]n Switzerland	they give you a cognac.
	Here	they give you tea and bikkies. (Eggins 1994: 272)
	THEMES	RHEMES

Another important aspect of thematic analysis is that thematic choices can be analysed at several levels of linguistic structure. In clause complexes made up of a main and a hypotactic clause such as (10), the analyst can describe the thematic structure of both (i) the clause complex, and (ii) the clauses that constitute it. In this case, as the hypotactic clause precedes the main clause, it is analysed as the topical Theme of the clause complex as a whole. Each clause within the complex can then be assigned its own Theme and Rheme. When there are more than one independent clause in the complex, however, the clauses are analysed for thematic structure independently of each other, as in (11). Traditionally, the basic unit of thematic analysis is the so-called 'independent conjoinable clause complex' (or T-unit), defined as "an independent clause together with all hypotactically related clauses which are dependent on it" (Fries 1994: 229).

(10)	As the universe	expanded	the temperature of the radiation	decreased. (Thompson 2014: 160)
	THEME		RHEME	
	THEME 1 (TEXTUAL + TOPICAL)	RHEME 1	THEME 2	RHEME 2

(11)	You	get a litre of milk	and it		stands this tall. (Eggins 1994: 291)
	THEME 1	RHEME 1	THEME 2 (TEXTUAL + TOPICAL)		RHEME 2

4.2.1.3   *The Rheme: The* parent pauvre *of thematic structure*

Whereas the Theme has been the focus of a large amount of research, so far the Rheme has tended to remain somewhat neglected in Systemic Functional research. Symptomatic of this is perhaps the fact that the Rheme is most frequently defined in negative terms with respect to the Theme. According to Bloor and Bloor (2004: 288), for example, the "Rheme is that part of the clause which is *not* the Theme" [my italics]. Similarly, for Thompson (2014: 147), the Theme is defined as the first constituent of the clause, whereas "*all the rest* of the clause is *simply* labelled the Rheme" [my italics]. In addition, as opposed to the Theme – within

which researchers have identified several subparts – the Rheme is generally viewed as one single, undifferentiated entity, and very few subdivisions have been made within it, although it usually covers a sizeable part of the clause. The lack of theoretical elaboration of the Rheme is discussed in rather vivid terms by Fawcett (2000). One major objection formulated by Fawcett concerning the SFL view of the Rheme is that it is not directly conflatable with the functional categories at other levels of metafunctional analysis. Yet, according to Fawcett (2000: 124), one ambition of SFL is to provide a representation of each clause as a set of functional elements across different metafunctions, which can then be mapped onto each other in order to obtain one single, coherent representation of that clause. In other words, according to Halliday (1977: 128, quoted in Fawcett 2000: 124):

> [I]t is the function of the lexicogrammar to map the structures onto one another so as to form a single integrated structure that represents all components [= 'components' of the grammar, in the sense of 'metafunctions'] simultaneously.

The problem with the Rheme is that it extends over more than one element in the other metafunctions. For instance, in Figure 8 (based on an example from Thompson 2014: 108), the Rheme in the textual metafunction is not coterminous with a single element in the experiential metafunction: instead, it encompasses the process along with two of its participants. Yet, as explained by Fawcett (2000: 127), SFL does not provide any indications as to how to conflate elements that are not coterminous across metafunctions.

EXAMPLE	I	explained	to her	what it meant
EXPERIENTIAL METAFUNCTION	Sayer (participant 1)	Process: verbal	Receiver (participant 2)	Verbiage (participant 3)
TEXTUAL METAFUNCTION	Theme	Rheme		

Figure 8. Mapping of the experiential and textual metafunctions

More fundamentally, in Fawcett's view, the lack of distinctions within the Rheme is also problematic because it gives the misleading impression that the Rheme constitutes one large element performing a single, coherent function whereas, for him, "there is not in fact a single 'functional element' called 'Rheme' that corresponds to several other clause elements" (ibid.: 132). Fawcett (ibid.: 138) goes as far as to suggest that the Rheme has in fact been put forward as a 'pseudo-element' intended to fill a gap in thematic structure, in order to give the illusion that every part of every clause can be accounted for through the prism of the three metafunctions. For him, however, the Rheme "do[es] not contribute in any way to the functional structure of the clause" (ibid.: 136). Accordingly, he advocates that the box which

is commonly labelled 'Rheme' in SFL analyses of thematic structure be simply left blank, as it does not correspond to any true function in the grammar (Fawcett 2000: 132; Fawcett 2007: 20).

Apart from Fawcett, a few other researchers have looked at the Rheme in a perhaps more positive light and have undertaken to provide a more detailed account of this unit than the one commonly found in SFL. However, research of this type is strikingly scarce[10]: I was only able to find three researchers seeking to make further distinctions within the Rheme. A first account is provided by Fries (1994). Starting from the observation that the Rheme as defined by Halliday is "too inclusive" (ibid.: 234), Fries insists on the importance of coming up with a rhematic equivalent to the notion of Theme. He coins the notion of 'N-Rheme' to refer to a subpart of the Rheme that corresponds to the last experiential constituent in the clause. Thus, just as the Theme refers to the initial part of the clause, the N-Rheme corresponds to the final element of the message. Functionally, the N-Rheme is claimed to constitute "the newsworthy part of the clause, that is, the part of the clause that the writer wants the reader to remember" (ibid.). The concept of N-Rheme is exemplified in (12), which is extracted from a fundraising letter. The purpose of the letter is to convince its readers to contribute to a project aiming to sensitise local politicians to the level of stress caused by population growth in large American cities. The N-Rheme is underlined in the example. By choosing to place this embedded clause in N-Rhematic position, the writer draws the potential contributor's attention to the stakes of the project advertised in the letter (i.e. exerting a positive influence on decisions which drastically affect the quality of the reader's life). Alternatively, the writer could have chosen to structure the information in the way exemplified in (13), thus laying emphasis on the frequency of the phenomenon at hand (i.e. every day). The N-Rheme is therefore assigned an important function as the carrier of the main informational focus of the message. It should be noted that the N-Rheme has been the focus of a number of later studies (see e.g. Cummings 2009; Herriman 2011; Elgemark 2017), which goes to suggest that Fries's elaboration of the Rheme filled a true gap in research on thematic structure.

(12)   Every day decisions are being made by local officials in our communities
       <u>that could drastically influence the quality of our lives</u>.   (Fries 1994: 243)

(13)   Decisions that could drastically influence the quality of our lives are being
       made by local officials in our communities <u>every day</u>.   (ibid.: 244)

Another relevant distinction within the Rheme is the one suggested by Morel and Danon-Boileau (1998). Based on the analysis of spoken French, Morel and

---

10.   The overview that follows is largely reproduced from Dupont 2015.

Danon-Boileau put forward the concept of 'post-rheme' ('*post-rhème*') to refer to an optional intonational constituent which consists of a group of syllables produced with a flat intonation, a low pitch and a reduced intensity at the end of an oral paragraph (ibid.: 28). Thus, Morel and Danon-Boileau's post-rheme is fully specific to spoken language. Semantically, the post-rheme may fulfil either of the following two functions: (i) express epistemic modality, as in (14) – where the post-rheme is underlined; or (ii) co-refer with a pronoun expressed earlier in the clause, as in (15) (ibid.).

(14)   En fait il y avait de la couleur ça amène un côté cru <u>moi je trouve</u> [Actually there were colours, it makes it look somewhat raw, <u>I think</u>].

(Morel & Danon-Boileau 1998: 29)

(15)   Mais elle est vieille <u>sa filleule</u> [But she's old, <u>his goddaughter</u>].   (ibid.)

Finally, Taglicht (1984) suggests that the notion of marked Theme as discussed by Halliday should be complemented with that of 'marked Rheme'. He defines the marked Rheme as a sentence-final constituent that is "detach[ed] from the element or elements with which it is contiguous in the unmarked sequence" (ibid.: 20). This detached constituent may be of three different types: (i) an end-shifted subject, such as *his face* in (16); (ii) an end-shifted constituent of the predicate, as in (17), where the direct object of the verb *found* (viz. *a number of interesting points*) is separated from it by a locative adjunct; or (iii) "a final item separated by a 'partition' from the item that would precede if it were part of the unmarked rheme" (ibid.: 23).[11] Interestingly, conjunctive markers are among the elements which can function as partitions between a marked rheme and the rest of the sentence, as in Example (18).

(16)   Even more terrifying was <u>his face</u>.   (Taglicht 1984: 23)

(17)   He had found in this article <u>a number of interesting points</u>.   (ibid.: 24)

(18)   They are returning, however, <u>to England</u>.   (ibid.: 25)

All in all, as can be seen from this brief overview, the rare researchers who have paid closer attention to the Rheme have tended to focus on the final part of the Rheme, leaving medial constituents to be virtually unaccounted for within the

---

11.   A partition can be defined as a textually parenthetical, syntactically extraneous element interrupting the flow of syntactically dependent elements of the clause (Taglicht 1984: 22). Partitions have frequently been referred to as 'parentheticals'.

system of thematic structure.[12] Nevertheless, in addition to the researchers who have explicitly added to the body of research on the Rheme, other scholars have made some distinctions with respect to the Theme which are in fact relevant for the study of the Rheme. Hannay and Gómez-González's (2012: 99) 'thematic parentheticals', for example, are defined as "parenthetical expressions, irrespective of form, which are marked off typographically and occur immediately after, and are triggered by, a thematic element". Thematic parentheticals are exemplified – and underlined – in (19) and (20). As shown in Example (20), conjunctive markers can take on the role of thematic parentheticals which, in fact, seem to bear some resemblance to Taglicht's (1984) partitions.

(19)   Recently, <u>particularly with respect to the Internet</u>, research into the possibility that some people are spending excessive amounts of time, and sometimes money, on computing activities has increased.

(Hannay & Gómez-González 2012: 100)

(20)   By the early 1970s, <u>however</u>, this attitude was changing and Sir Robert Mark […] promised to do away with corruption in the force.   (ibid.: 104)

Because they are triggered by a thematic element, Hannay and Gómez-González consider thematic parentheticals to be part of the Theme – even though the authors acknowledge that some of them are Rheme-oriented, viz. perform the function of providing additional information on the subsequent, rhematic part of the message. Following Halliday's definition of Theme, however, strictly speaking these elements are rhematic, as they occur after the topical Theme (i.e. temporal adjuncts functioning as marked Themes, in the two examples above).

In conclusion, distinctions pertaining to the Rheme, though infrequent, are available. Nevertheless, they are scattered across the literature: there seems to be no consensus on the kinds of subdivisions to be made within the Rheme or their

---

12.   It is worth noting that further distinctions within the Rheme are found in Functional Sentence Perspective (FSP), a theory developed by the Prague School of linguistics which has been a major inspiration for Systemic Functional linguists in the elaboration of the systems of information and thematic structure. In FSP, clauses are divided into theme and 'non-theme', which is itself made up of (i) transitional elements and (ii) a 'rheme proper'. FSP makes distinctions within the non-theme by resorting to the Given/New distinction, with different parts of the non-theme corresponding to different degrees of newness of the information, or so-called 'communicative dynamism' (see e.g. Firbas 1992; 1995; Hajičová 1994; see also Hartnett 1995, who focuses on the slot which comes right after the Theme). The FSP perspective on thematic structure, however, is hardly comparable to the SFL one, notably because (i) no clear separation is made between the systems of Theme/Rheme and Given/New; and (ii) as opposed to SFL, FSP does not conceive of Theme and Rheme as position-bound notions (see e.g. Firbas 1995: 214).

denominations, and within the small community of researchers interested in the Rheme, scholars make little or no reference to each other's work. Yet, as will be shown later in this chapter, making distinctions within the Rheme is important for the study of some syntactic phenomena, including conjunctive adjunct placement.

4.2.1.4 Controversy over thematic structure: Theme/Rheme boundary and cross-linguistic validity

At the end of the previous section, we saw that Hannay and Gómez-González (2012) adopt a slightly extended definition of Theme, which includes any parenthetical expression following a thematic element and triggered by it, instead of stopping at the first experiential constituent in the clause. This raises an important aspect of research into thematic structure, viz. the fact that the delimitation of the Theme has aroused a great deal of controversy among Systemic Functional linguists (see e.g. Berry 1996: 29–31; Butler 2003: 127–145 for overviews). The main sticking point in this respect has concerned the status of the subject of clauses starting with a marked Theme, such as *most Americans* in Example (21).

(21)   In 1800 <u>most Americans</u> were farmers by occupation.

(Thompson & Hunston 2008: 55)

In traditional thematic analysis, the Theme would only include the temporal adjunct *In 1800*, which constitutes the first experiential element in the clause. However, other researchers have advocated setting a fixed Theme/Rheme boundary after the grammatical subject of the clause, irrespective of whether it is preceded by other experiential elements. For them, then, the Theme in Example (21) would extend up to and including *most Americans* (see e.g. Enkvist 1973; Downing 1991; Caffarel 2006). The argument advanced in favour of such an analysis is that, as the typical carriers of the topic of the clause, subjects constitute the true starting point of the message, and thus should form an integral part of the Theme (Thompson & Hunston 2008: 57).[13] Marked Themes, on the other hand, do not convey the topic of the message, but merely act as situational frameworks which orient the message in a certain context. As such they do not saturate the experiential potential of the Theme (Downing 1991; Rossette 2009: 18). In other words, as Ravelli (1995: 227, quoted in Thompson & Hunston 2008: 58) argues, "until a potential Subject element is confirmed, the Theme analysis is still relevant, as the message is not yet fully 'off the ground'".

---

13.   This debate partly stems from the fact that, as already explained above, Halliday's initial definition of Theme was twofold: Theme was defined both as the *starting point* of the message, and as the element determining what the message was going to be *about*. This initial view is still reflected in the term '*topical* Theme'.

Further extensions of the Theme have been put forward as well. Berry (1996), for example, undertakes an analysis of the distribution of 'typically thematic meanings' throughout the clause (which she defines as the meanings which are prioritised by the writer, and correspond to his/her main concerns). Based on the analysis of three texts produced by the University of Nottingham for prospective and new students in the Humanities, she suggests that the boundary between Theme and Rheme is best set after the lexical verb, as in (22). Gómez-González's (2001: 329–346) 'Extended Multiple Theme' (on which Hannay and Gómez-González's thematic parentheticals are based) expands the traditional definition of multiple Theme to also include any interpersonal or textual elements *following* the topical Theme, as in (23), along with any parenthetical elements modifying and directly following the topical Theme, as in (24). Matthiessen (1992), for his part, finds it misleading to set a clear boundary between the Theme and the Rheme: instead, thematic structure is to be viewed as a continuum, with thematic prominence decreasing as we move towards the end of the clause.

(22)	With the consent of the Head of department a candidate may substitute	a dissertation in English […] in of 20 credits worth of Level B or Level 3 modules. (Berry 1996: 45)
	THEME	RHEME

(23)	West Morland for example	became particularly passionate when talking about the influence of television. (Gómez-González 2001: 330)
	EXTENDED MULTIPLE THEME	RHEME

(24)	Perhaps she, though so small, of so quick perishing beauty,	is nonetheless part of His immortal dream. (ibid.: 331)
	EXTENDED MULTIPLE THEME	RHEME

In the present book, the traditional view of the Theme, extending up to and including the first experiential element in the clause, is adopted. Two main reasons justify such an approach. The first one pertains to the central SFL idea of choice, which was already referred to at several points in this chapter. The underlying principle behind the 'one topical Theme per clause' rule is that the element functioning as topical Theme is significant precisely by virtue of being the result of a choice on the part of the writer or speaker. Accordingly, one of the main interests of thematic analysis is that it sheds light on what type of elements are chosen, from among a range of options, as the point of departure for a given message in a specific context (Eggins 1994: 284). As explained by Thompson and Hunston (2008: 58),

establishing the subject as an obligatory component of the Theme nullifies this fundamental aspect of the concept of Theme. The second argument against an extended definition of the Theme is that equating subject and topical Theme blurs the boundaries between the metafunctions (ibid.): whereas the Theme is a textual notion, in SFL the subject pertains to the interpersonal metafunction of language (see Halliday & Matthiessen 2004: 117–119 for more details). This means that, even though there is a natural affinity between the two constructs, Theme and subject make distinct contributions to the meaning of the message (with the subject acting as the "resting point of the argument", i.e. the element with respect to which the content of the clause is claimed to have validity, as opposed to the Theme, which functions as the starting point of the message, see Thompson & Hunston 2008: 58). As a result, the level of detail of the analysis incurs an appreciable loss if subject and Theme are simply merged.

A final issue to be discussed in this section has to do with the cross-linguistic validity of the definition of Theme. More particularly, the question raised here is whether the same criteria of identification of the Theme can be applied to both English and French. Although, according to SFL, every language has at its disposal a set of resources to signal what elements should be viewed as the point of departure of the message, the way this function is *realised* may vary across languages (Halliday & Matthiessen 2004: 64; Fries 1995: 15). In line with most SFL theories, the system of thematic structure was developed primarily in relation to the English language; therefore, the Theme is most commonly associated with the clause-initial position. However, in some languages the function of Theme is realised by other means than word order, such as particles or affixes (e.g. in Japanese or Tagalog, see Fries 1995: 15). In other languages, the Theme does correspond to the clause-initial position, but different word order principles still call into question the direct applicability of the rules of identification and characterisation of the Theme established on the basis of English. In most Germanic languages, for example, the 'verb-second' (V2) constraint, which only allows for a single major constituent to occur before the finite verb of the clause, may induce differences with English – a predominantly 'subject-verb-object' (SVO) language – as to the degree of markedness associated with certain thematic choices. In languages such as Norwegian, Swedish or German, for instance, the V2 rule makes it more common – and hence less marked – than in English for non-subject elements to occur in thematic positions (see e.g. Hasselgård 2004a: 188; Altenberg 1998: 117; Teich 2003: 123). Fortunately, with respect to French, the adaptation of the definition of Theme developed for English is fairly straightforward. French word order operates according to the same broad SVO principles as English. As a result, the definition of the French Theme

is very similar to the definition of the Theme in English (Cummings 2009: 45).[14] In both languages, the Theme corresponds to the clause-initial position, and its unmarked realisation in declarative clauses is the subject (Caffarel 2006; Rossette 2009: 11; Cummings 2009: 45; Banks 2017: 45).[15] It is striking, however, that among the few researchers who have discussed French thematic structure, a majority adopt an extended approach to the Theme, considering that it should always stretch up to and including the subject of the clause (e.g. Caffarel 2006; Cummings 2009; Rossette 2009; see also above on the debate concerning the delimitation of the Theme). Thus, for these researchers, French Themes may either (i) contain a single, unmarked topical Theme; or (ii) consist of a marked topical Theme, followed by an unmarked topical Theme. In accordance with the arguments put forward earlier in this section, however, I have decided to adopt the Hallidayan view of Theme, which stipulates that the Theme stops after the first experiential element of the clause, in both English and French.

4.2.2 Thematic structure and conjunctive markers

Whereas the aim of Section 4.2.1 was to provide a general description of the system of thematic structure, this section explores the connections between thematic structure and cohesive conjunction. There is a close relationship between these two linguistic systems: conjunctive markers have commonly been described as standing in close association with the Theme of the clause or clause complex. In fact, Halliday and Matthiessen (2004: 83) have described conjunctive markers as

---

14. The major difference with English is that in French, the unmarked Theme may vary within a particular mood (Caffarel 2006: 181). The clearest example pertains to interrogative clauses, which can take two main forms in French, viz. (i) *Que manges-tu ?* or (ii) *Tu manges quoi ?* In such cases, the degree of markedness of each option will mainly depend on the context of utterance of the interrogative clause (e.g. formal or informal).

15. Note that within the broad SVO structure, there are some differences in the preferred word order patterns of the two languages. For example, whereas English clauses tend to start with the subject, in French it is more frequent for subjects to be preceded by a range of optional elements such as adjuncts – as in: *Dehors, la blancheur fondait rapidement dans les champs […],* translated into English as: *The whiteness was burning rapidly off the field outside […]* – but also dislocated subjects or objects, participle clauses, etc. (Chuquet & Paillard 2017: 125–126; see also Chuquet & Paillard 1987: 156; Vinay & Darbelnet 1995: 214; Trévise 1986). However, such word order differences do not seem to significantly affect the degree to which the English system of thematic structure can be adapted to the French language (e.g. even though they seem to be more frequent in French than in English, fronted adjuncts and dislocated subjects/objects are still analysed as marked Themes by French Systemic Functional linguists; see e.g. Caffarel 2006: 172 on this).

being either inherently, or characteristically thematic, depending on their category. Both coordinators and subordinators can be said to be inherently thematic, in the sense that they are fixed at clause boundary, and thus may *only* occur in clause-initial (or thematic) position (Halliday & Matthiessen 2004: 83; Quirk et al. 1985: 921; see also Chapter 2 on this). By contrast, conjunctive adjuncts display some degree of syntactic mobility. They may occur not only sentence-initially – as in (25), but also medially and finally, as in (26) and (27), respectively (Biber et al. 1999: 890–891). The mobility of these markers makes them particularly interesting and will be the object of Chapter 8.

(25)   After the Bay of Pigs fiasco, he said ruefully, "it would have been better if we had left it to James Bond". **On the other hand**, his reputed attempts to get Castro to extinguish himself with either an exploding cigar or a poison pen may have owed all too much to Bond.   (Halliday & Matthiessen 2004: 544)

(26)   These characteristics, **however**, are dependent on other variables such as the conditions of pressure and temperature at entry to the compressor and the physical properties of the working fluid.   (Biber et al. 1999: 892)

(27)   Mr Straw played down the significance of the achievement. He should not be so modest. There are important caveats, **nevertheless**.   (Dupont 2015)

Despite their syntactic mobility, however, conjunctive adjuncts have frequently been said to display a strong preference for the thematic (i.e. initial) position. This tendency has been claimed to originate from their intrinsic linking function: as signals of the relationship between a given textual unit and the preceding discourse, it makes sense that they should occur as early as possible in the clause, to facilitate discourse processing by the reader (Altenberg 2006: 11; Lenker 2014: 15). Therefore, according to Halliday and Matthiessen (2004: 83), even though they may occur at various points in the clause, conjunctive adjuncts are 'natural Themes':

> [Conjunctive adjuncts] are what we might call *characteristically thematic* […]. [T]hey are […] *natural Themes*: if the speaker, or writer, is making explicit the way the clause relates to the surrounding discourse […] it is *natural* to set up such expressions as the point of departure [my italics].

Equivalent statements are also widespread in the non-SFL literature, at least with respect to English: the initial position has been referred to successively as the 'normal', 'unmarked' or even the 'default' position for English conjunctive adjuncts. According to Quirk et al. (1972), for instance: "[t]he normal position for most conjuncts is I [initial] […]. M [medial] positions are rare for most conjuncts, and E [end] rarer still". Likewise, for Biber et al. (1999: 891) the "initial position can […] be considered as the unmarked position for linking adverbials" (see also Quirk et al. 1985: 643; Leech & Svartvik 2002: 139; Lenker 2010: 198 for similar

statements). Thus, there is some insistence, in the literature, on a strong associa-
tion between conjunctive adjuncts and the thematic portion of the clause. In SFL,
this is epitomised by the notion of 'textual Theme': the relationship between con-
junctive markers and the Theme is such a close one that Halliday found it useful
to create a specific slot for markers of conjunction within the Theme. By contrast,
no equivalent theoretical concept is available for CAs occurring later in the clause,
viz. within the Rheme. Importantly, this does not mean that SFL has completely
overlooked the potential of CAs for syntactic mobility: some Systemic Functional
linguists – including Halliday himself – have shown awareness of the fact that other
positions than thematic are possible for CAs. In the first edition of his *Introduction
to Functional* Grammar, for instance, Halliday (1985: 81) states the following:

> Conjunctive Adjuncts tend to occur at points in the clause which are significant
> for textual organization, which means at some boundary or other: (i) clause ini-
> tial, as (part of) the textual theme; (ii) clause final, as afterthought; (iii) between
> Theme and Rheme; (iv) between Mood and Residue.[16]

Hartnett (1995) stresses that it is relatively common for CAs to occur in what
she calls the 'Pit after the Theme', i.e. the position that immediately follows the
Theme.[17] Likewise, Thompson (2014: 163) explains that, although CAs are "natu-
ral starting points" for the clause, they do not *have to* be thematic, as the writer or
speaker can also choose to place them at the Theme-Rheme boundary, or "even
later in the Rheme". Such an awareness has tended to remain limited to a few com-
ments made in passing, however, and very little SFL research has undertaken to
explore the rhematic uses of CAs in further detail.

Yet, there is evidence in the literature that other positions than the Theme are
significant for conjunctive adjuncts, which highlights the need to qualify, at least
to some extent, the strength of the relationship between CA placement and the
Theme. Firstly, several studies which have adopted a contrastive approach to CA
placement have shown that the degree to which CAs tend to be placed themati-
cally varies, sometimes significantly, across languages. For example, based on an
English-Swedish corpus made up of both fiction and non-fiction texts, Altenberg
(1998; 2006) demonstrates that, whereas English tends to use a majority of con-
junctive adjuncts in initial position, in Swedish the most frequent slot is the medial
position. Looking at the translation of English multiple Themes into German and

---

16.   In SFL, the Mood and the Residue belong to the interpersonal metafunction. The Mood
is the part of the clause that consists of (i) the subject and (ii) the part of the verbal group
which expresses tense or modality (Bloor & Bloor 2004: 284). The Residue corresponds to the
rest of the clause (i.e. what is not the Mood).

17.   Hartnett's (1995) study is situated at the boundary between SFL and FSP.

Norwegian, Hasselgård (2004b: 82) observes that a notable number of textual Themes in English are transferred to the Rheme of the Norwegian and German sentences, probably as a result of the V2 constraint which operates in these two languages. Hasselgård's results are in line with Kunz and Lapshinova-Koltunski (2014), who report that initial conjunctive adjuncts are about twice as frequent in English as in German written texts (see also Chapter 8 on this).

It is significant that, although some of these contrastive studies adopt an SFL approach to conjunctive markers, they usually resort to purely syntactic labels (e.g. initial, medial, final positions) to describe CA placement, making little or no reference to the functional notions of Theme and Rheme. This phenomenon is likely due to the fact that the SFL system of thematic structure – i.e. the system that is supposedly concerned with the linear ordering of constituents in the clause, see Eggins (1994: 273) – does not provide any detailed resources for describing the position of elements occurring later than the topical Theme: as explained in the previous section, SFL represents the rhematic part of the clause as one undifferentiated unit. As already discussed earlier, the system of thematic structure was elaborated primarily in relation to the English language. Given the apparent predominance of initial CAs in English, it is probably not very surprising that the relationship between Themes and conjunctive adjuncts has been paid such considerable attention, whereas the rhematic uses of these items have been granted little consideration. In view of the results obtained when analysing CA placement in a number of other languages, however, it appears important to pay closer attention to the different positional options available within the Rheme. In fact, one may wonder whether the textual Theme would have received so much focus, had thematic structure been developed on the basis of another language. It is perhaps conceivable that, faced with another linguistic system, Halliday might have developed (a) rhematic equivalent(s) to his textual Theme.

Although no study has yet undertaken to systematically compare the placement patterns of CAs in English and French on a large empirical basis, a survey of the monolingual literature on CA placement in the two languages suggests that similar differences in use of thematic CAs as the ones uncovered for the English-Swedish or English-German language pairs are to be expected. Whereas, as explained earlier, descriptions of CA placement in English strongly emphasise the tendency of the markers to occur initially, in French the claims pertaining to CA placement are somewhat more nuanced: firstly, the emphasis is usually laid primarily on the mobility of conjunctive adjuncts in the clause (e.g. Rubattel 1982: 59; Hawkins & Towell 2001: 123–124); secondly, when reference is actually made to the preferred positions of conjunctive adjuncts, the tendency of CAs to occur medially – and, more particularly, within the verb phrase – as well as initially is highlighted. According to Grevisse and Goosse (2011: 1211), for example,

in sentences whose verb phrase contains an auxiliary, "adverbs that signal a logical relation are most frequently placed between the auxiliary and the participle" [my translation] (cf. also Csüry 2006: 111; see Chapter 8 for more details). Therefore, whereas we saw in Section 4.2.1.4 that the same definition of Theme can be used for both French and English, it is likely that the description of French CA placement will require further development of the notion of Rheme.

Even with respect to English, a few studies have provided evidence which has somewhat qualified the claims generally made on the placement patterns of CAs. Greenbaum (1969: 78–80), for instance, shows that the marked tendency of CAs to occur sentence-initially does not seem to apply with equal force in all communicative situations. Greenbaum analyses the position of English conjunctive adjuncts in a corpus of about 100,000 words spread across eight different registers. He observes that, while the initial position accounts for about 75% of the total number of occurrences of conjunctive adjuncts in his corpus, there is noticeable variation in placement across the subcomponents of his data set. For example, a majority of the final markers in his data occur in the discussion and conversation subsections of the corpus, whereas most of the medial conjunctive adjuncts come from scientific texts (cf. also Biber et al. 1999: 891; Conrad 1999 for cross-register analyses of CA placement). In addition, some researchers have uncovered significant variation in placement between English conjunctive adjuncts. Altenberg (2006: 15), for example, finds "clear evidence of individual positional 'profiles'" among English conjunctive adjuncts. He demonstrates that, alongside the markers which display a clear tendency to occur initially (e.g. *besides, in other words*), some conjunctive adjuncts show a definite preference for the medial (e.g. *therefore, accordingly*) or the final (e.g. *though, then*) slots. Finally, Altenberg (2006) and Lenker (2010; 2011; 2014) have stressed that, although non-initial positions are not the most frequent slots for English CAs, they are nonetheless significant by virtue of the rhetorical effects that they can produce. Altenberg (2006) and Lenker (2010; 2011; 2014) demonstrate that non-initial CAs, and more particularly CAs occurring in sentence-medial positions, usually take on further roles (e.g. focus, emphasis) in addition to their inherent linking function. Thus, some scholars (mostly outside SFL) have shown that even in English, it may be of interest to pay closer attention to the placement patterns of CAs occurring later than the Theme. Nevertheless, this more 'nuanced' type of research into English CA placement has tended to remain somewhat scattered and piecemeal. Further work is therefore needed to get a better sense of the respective influence of lexical and/or register factors on the placement of English conjunctive adjuncts. These questions will be the focus of Chapter 8.

In summary, whereas in both Systemic Functional Linguistics and English linguistics in general, conjunctive adjuncts have predominantly been associated

with the initial or thematic position, some researchers have provided evidence that other positions are also worth investigating, by taking account of variation in placement across languages, registers or lexical items. Such studies have acted as prompts for me to pay closer attention to patterns of CA placement after the topical Theme. Unlike previous research into CA placement, and in line with the theoretical framework adopted in this study, I intend to analyse CA placement by relying fully on the SFL system of thematic structure, i.e. by always describing the position of CAs of contrast with respect to the notions of Theme and Rheme. Yet, as I hope to have made clear in this chapter, whereas the SFL framework provides a very detailed description of the Theme, the Rheme has remained a neglected object of study. Therefore, one 'meta-objective' of the analyses of CA placement carried out in this study will be to put forward subdivisions within the Rheme which will allow for a detailed and accurate description of CA placement in both English and French authentic corpus data. The decision to make further distinctions within the Rheme is especially important in view of the cross-linguistic approach adopted here: relying on the coarse positional distinctions currently available in the system of thematic structure would run the risk of overlooking some relevant differences between the languages, especially as the SFL framework adopted here is character-ised by a bias towards the English language.

4.3   Conclusion

This last theoretical chapter has revolved around two main axes. First, it pre-sented the main benefits that can be gained from the approach adopted in this book, viz. a combined Systemic Functional and corpus approach. The two lin-guistic frameworks were shown to display both similarities and divergences in the ways in which they address linguistic description. On the one hand, SFL and CL were shown to share a number of concerns, including the importance placed on authentic language, the significance attached to quantifications for language theorisation, or the attention paid to the influence of the communicative situa-tion on the frequency and use of linguistic phenomena. On the other hand, it was explained that Systemic Functional and corpus linguists take a markedly different stance on these common areas of interest: whereas SFL scholars adopt a predomi-nantly theory-oriented approach to language, corpus linguists are primarily inter-ested in providing the most accurate description possible of the usage patterns observed in authentic linguistic data (while developing appropriate methodologi-cal tools to do so). Instead of representing difficulties in combining SFL and CL, however, these diverging approaches to linguistic description were shown to be complementary, with the strengths of SFL compensating for the shortcomings of

CL, and vice-versa. A dialectical relationship between SFL and CL was therefore advocated, where the corpus data should serve to refine the SFL theory, while SFL should be used to give shape to and interpret the tendencies emerging from large corpora. The first part of this chapter ended with an overview of some of the recent research combining the SFL and the corpus approaches. It was shown that, even though the synergy between SFL and CL is well underway, there is still scope for research helping to strengthen the ties between them. This is one of the objectives of this study.

The second part of this chapter focused on one specific area of SFL which I intend to further elaborate through corpus methods, viz. the textual metafunction and, more particularly, the system of thematic structure. The two main constituents of thematic structure, viz. the Theme and the Rheme, were defined, and it was shown that, whereas the Theme has attracted a lot of attention in SFL, the Rheme has remained an understudied theoretical construct. The last section of the chapter was devoted to a discussion of the close relationship between thematic structure and conjunctive markers. It was shown that in both Systemic Functional Linguistics and English linguistics in general, emphasis is usually laid on the tendency of conjunctive adjuncts to occur clause-initially, i.e. within the thematic part of the clause. However, some research has drawn attention to the significance of rhematic positions for conjunctive adjuncts, including when looking at other language systems than English. Nevertheless, as the Rheme has tended to be portrayed as a single entity, thematic structure in its current form is not sufficiently detailed to provide an accurate picture of CA placement, whether in English or in other languages. Therefore, one objective of this book will be to identify the components of the Rheme which are relevant for the cross-linguistic description of the placement patterns of conjunctive adjuncts of contrast.

Data and methodology

One of the central objectives of this study is to revisit some of the intuition-based contrastive claims on English and French conjunctive marker usage on a large empirical foundation. This presupposes the choice of appropriate corpus resources, as well as efficient methods to analyse them. These choices are the object of the present chapter. Section 5.1 presents the corpus data used in this study, while Section 5.2 describes the methodological steps taken to analyse it.

5.1 Data

5.1.1 Comparable or translation data?

In Chapter 3, two main types of corpus resources were identified for contrastive linguistics: (i) comparable corpora, made up of original texts in the two (or more) languages compared, matched by criteria such as text type, time of composition, target audience, etc.; and (ii) translation corpora, which consist of original texts in one language along with their translations into the other language(s). Both types of corpus were shown to have strengths as well as weaknesses: for example, while comparable corpora have the advantage of giving access to original, unmediated data (viz. data that is not influenced by the translation process) in the languages compared, it raises a number of issues pertaining to comparability, both in terms of the types of texts included in the corpus, and the linguistic units to be contrasted across languages. Such comparability issues do not arise with translation corpora: on the one hand, as they are translations of each other, the texts that they contain are in principle equivalent from a semantic, pragmatic and textual point of view. On the other hand, translation data provides the researcher with a ready-made *tertium comparationis*: the linguistic units to be compared are directly accessible through the relation of translation equivalence that unites them. However, the main limitation of translation corpora is that the data which they give access to is likely to diverge from original language by displaying traces of influence from the source language, as well as a range of properties induced by the translation process (e.g. a tendency to be more explicit than corresponding original language,

or a somewhat exaggerated conformity to the norms of the target language). As a result, the frequency and distribution of linguistic phenomena are likely to differ in translated as compared to original texts (Lewis 2009: 190). As pointed out by Johansson (2007: 21), there is not *one* good solution with respect to the choice of an appropriate data set for contrastive research: rather, "the choice of a [corpus] model is dependent upon the research question and the object of the study".

Conjunctive markers, in particular, have been shown to be especially sensitive to the influence of the translation process. As discussed in Chapter 2, they are often optional, which means that the relations that they convey can often be inferred even in their absence. One consequence of the optionality of CMs is that they are highly volatile in translation, i.e. they are frequently added, removed, or undergo semantic and syntactic changes in the transition from one language to another (e.g. Halverson 2004: 570; Cartoni et al. 2011: 79; Dupont & Zufferey 2017: 271). While some of these shifts may be triggered by the conventions of the target language, research has shown that this factor is only part of the picture, and that other phenomena, such as translation-related explicitation for example, are also likely to come into play. In this respect, some studies have uncovered significant quantitative differences in frequency of CMs between comparable corpora of original and translated texts in the same language. Xiao and Yue (2009: 257), for example, compare the frequency of Chinese conjunctive markers in a corpus of original and translated fiction (with the translated subcorpus containing texts translated mostly from English, but also from a range of other languages, including French, German, Spanish). They find that Chinese conjunctive markers are highly significantly more frequent in the translated than in the original texts, and that the range of markers (i.e. CM types) found in the translated data is nearly twice as varied as in the original data. Xiao (2010a: 23–24) also uncovers a significantly higher overall frequency of conjunctions in a genre-balanced corpus of translated Chinese (including the same range of source languages as in Xiao and Yue 2009) as compared to a comparable corpus of original texts. He notes that the overuse of conjunctions in translation mostly pertains to high-frequency conjunctions, whereas low-frequency conjunctions are more frequent in the original than in the translated corpus. In a similar vein, Becher (2011) observes that in a bidirectional English-German translation corpus, additions of CMs in translation from English into German are not counterbalanced by an equivalent frequency of CM omissions in the other translation direction, which suggests that shifts in the use of CMs are not solely due to cross-linguistic differences between English and German (see also Mason 2001: 67 on this). Finally, Cartoni et al. (2011) compare original and translated texts in a bidirectional translation corpus of English and French parliamentary debates, using two distinct measures of comparison: (i) a distributional measure of lexical similarity and (ii) a comparison of CM frequencies.

While they find no significant differences between original and translated texts in terms of lexical distribution, with respect to the frequency of CMs they uncover marked differences between both original and translated English, and original and translated French (ibid.: 83). Comparing source and target texts, Cartoni et al. also report that overall, causal CMs are inserted much more frequently than they are removed during the process of translation, whatever the translation direction: in translation into English, translators added twice as many CMs as they removed, while in French, the ratio of added vs removed CMs rose to more than four to one (see also Øverås 1998 for similar results on English and Norwegian). A second stage of Cartoni et al.'s study aims to assess the impact of the source language on the translation of CMs. This part of the analysis is based on a corpus of French texts translated from four source languages (English, Italian, German and Spanish). Cartoni et al. (2011: 84) highlight significant differences in the distribution of CMs across corpora of French texts translated from these four source languages: they show that each source language appears to increase the frequency of one specific causal marker (e.g. in French translated from German, *car* is used twice as frequently as in the other subcorpora; *étant donné que* is twice as frequent in the texts translated from English as in the other texts, etc.).

All these results demonstrate that translated language does not behave in the same way as original language with respect to CM use. Firstly, the fact that translated language tends to display higher frequencies of CMs than original language, along with the fact that CMs are more often added than removed in translation, seems to indicate that CMs are especially prone to translation-inherent explicitation, i.e. explicitation that is triggered by the mere translation activity, instead of being traceable to differences between the languages involved. As a matter of fact, it is significant that, although explicitation in translation can pertain to a diversity of linguistic phenomena (e.g. additions of the optional reporting *that* in translation; see Laviosa 2009: 308; Mauranen 2008: 38–39 for more details), the first mention of the 'explicitation hypothesis' was made by Blum-Kulka (1986) in relation to cohesion.[1] In addition to translation-inherent explicitation, the use of

---

1. It should be noted that explicitation, and translation universals in general, have been strongly challenged in the field of translation studies, with many researchers arguing that it is too radical to refer to recurrent features of translated language as being 'universal' (see e.g. Becher 2011; see also Mauranen 2008: 34–36 for a brief summary of the debate). Accordingly, a number of researchers have put forward alternative terms to that of 'translation universals', among which 'translation laws', 'translation tendencies' (Mauranen 2008: 34), 'properties of translation' (Neumann 2013; Neumann 2014) or 'features of translated language' (Kruger & van Rooy 2012; Kruger 2017). It is beyond the scope of this study to further expand on these issues, however.

CMs in translated data also seems to reflect the influence of the source language, as attested by Cartoni et al.'s (2011) comparison of translated texts resulting from different source languages (see also Granger 2018 on the influence of the source language on the translation of CMs).

From all this it follows that researchers aiming to compare the use of CMs in two or more languages should pay very close attention to selecting the corpus that is most suitable for their research objectives. In the state-of-the-art survey presented in Chapter 3, it was shown that a large proportion of corpus-based contrastive studies on CMs have been concerned with assessing the degree of equivalence between markers across languages. A broad overview of the type of results obtained in such studies demonstrated the great value that translation data can have in granting access to the semantic specificities of individual markers and pinpointing very subtle usage differences between CMs, both within and between languages. For a more onomasiological study such as the present one, however, which aims to compare the general strategies of discourse organisation of two or more languages (e.g. frequency and distribution of explicit CMs, preferred types of CMs used in a specific context), resorting (exclusively) to translation data may prove inadequate, as it is sometimes difficult to know for certain whether the tendencies observed in the translated data should be attributed to genuine cross-linguistic differences, or result from the process of translation. Rather, for this type of study, comparable corpora of original texts emerge as a more reliable option. For this reason, all the analyses in the present book are carried out on comparable data. As mentioned above, such a methodological choice raises other issues, and more particularly that of comparability. When using comparable data, researchers must ensure the comparability not only (i) of the texts that are being analysed in each language, but also (ii) of the linguistic units investigated, viz. they need to establish a reliable *tertium comparationis* for their study. The latter issue was discussed at length in Chapter 2. The former issue, viz. text type comparability, is discussed in the next section.

Another alternative might have been to use both comparable and translation data to compare English and French CM usage. As made clear in Chapter 3, the strengths of one corpus type usually correspond to (and thus complement) the weaknesses of the other, so that a number of researchers have advocated combining the two types of resources: Johansson's (2007) 'bidirectional translation model', for example, is based on the use of a bidirectional translation corpus (e.g. French original texts translated into English, and vice-versa), whose original subparts can also be used as a comparable corpus. The problem with such corpora is that, due to the considerable difficulty of getting access to sentence-aligned translation data in both translation directions, they tend to be fairly small, and limited to a restricted number of (sometimes highly specific) language registers (see Chapter 3 for more

details). Yet, as CMs are not particularly frequent linguistic phenomena in written texts, they require to be investigated in fairly large amounts of data.

5.1.2   Description of the corpus data

The present study is based on a comparable corpus of English and French original texts including two subcomponents, each representing a different language register. The first subcorpus is the *Multilingual Corpus of Editorials* (Mult-Ed), which is made up of quality newspaper editorials in the two languages.[2] The second subcorpus corresponds to the bilingual subpart of the *Louvain Corpus of Research Articles* (LOCRA) and consists of English and French research articles in the Humanities.[3] Such a corpus design makes it possible to offer some register-sensitive cross-linguistic comparisons, in line with the recent insistence, in corpus-based contrastive linguistics, on the importance of taking register variation into account when formulating similarities and differences between languages (e.g. Lefer & Vogeleer 2014; see also Chapter 3; Section 3.1.3 for a discussion). This section provides a detailed description of each subcorpus, while laying special emphasis on the crucial question of data comparability, which is essential to ensure the validity of the cross-linguistic comparisons performed in Chapters 7 and 8.

A summary of the main features of each subcorpus is provided in Table 5. First of all, the table shows that the English and the French texts in each subcorpus were made as comparable as possible with respect to register. In Mult-Ed, the English and the French editorials alike come from quality newspapers, such as *The Independent, The Times* or *The Guardian* in English (as opposed to tabloids such as *The Sun* or *The Mirror*), and *Le Figaro, Libération* or *Le Monde* in French.[4] Note that each subpart of the Mult-Ed subcorpus is restricted to a single language variety, viz. British English and Metropolitan French, respectively. Likewise, both the English and the French subparts of the LOCRA subcorpus consist of research articles compiled from top-ranked journals in the Humanities. Furthermore, in both languages, the texts in the corpus are evenly distributed across the same five academic disciplines, viz. anthropology, education, political science, psychology and sociology.[5]

---

2.   <https://uclouvain.be/en/research-institutes/ilc/cecl/mult-ed.html> (12 April 2020).

3.   <https://uclouvain.be/en/research-institutes/ilc/cecl/locra.html> (12 April 2020).

4.   See Appendix 1 for a complete list of the newspapers and academic journals from which the texts were extracted.

5.   In LOCRA, all non-scientific citations (e.g. personal anecdotes in anthropology articles; interviews with students in the education papers; extracts of political speeches in the political science articles, etc.) were deleted from the data, so as to make sure that the corpus would only

As appears from Table 5, the corpus is also balanced with respect to the size of its subparts. All the word counts provided in the table were computed with the sixth version of the WordSmith Tools corpus analysis software (Scott 2012), using the tool's default count parameters for each language. Both the LOCRA and the Mult-Ed subcorpora contain about 2 million words per language, amounting to over 8 million words in total. Within each register, cross-linguistic comparability is also ensured in terms of both (i) the number of texts (viz. c. 300 texts per language in LOCRA; c. 4,000 texts per language in Mult-Ed) and (ii) the size of the texts included in the corpus (with an average of about 7,000 words per text in LOCRA; and 480 words per text in Mult-Ed).[6] Finally, for each register, the times of publication of the texts are also closely comparable between the English and French subparts of the corpus.

Table 5. Main features of the LOCRA and the Mult-Ed subcorpora

		ORIGINAL ENGLISH	ORIGINAL FRENCH
LOCRA	REGISTER	Research articles in the Humanities from top-ranked journals across 5 disciplines (viz. anthropology, education, political science, psychology and sociology)	Research articles in the Humanities from top-ranked journals across 5 disciplines (viz. anthropology, education, political science, psychology and sociology)
	TOTAL # OF WORDS	2,033,106	2,025,372
	TOTAL # OF TEXTS	271	313
	AVERAGE TEXT LENGTH	7,502 words	6,470 words
	PUBLICATION DATE	2007-2013	2006-2014

---

include academic discourse. The abstracts were also removed, as they are sometimes considered as an academic (sub)register in themselves: for example, they may stand on their own as independent textual units, and tend to follow their own discourse conventions (see e.g. Lores 2004; Cross & Oppenheim 2006). Due to the very short, self-contained nature of these texts, the use of conjunctive markers is likely to be very atypical in research abstracts, and it was thus deemed preferable to exclude them from the analysis.

6.   In this respect, it must nevertheless be noted that the length of the texts in all four subparts of the corpus (viz. LOCRA-EN, LOCRA-FR, Mult-Ed-EN and Mult-Ed-FR) varies very widely, viz. (i) between 2,136 and 19,841 words for LOCRA-EN (with a standard deviation of 2,416 words); (ii) between 1,304 and 16,492 words for LOCRA-FR (standard deviation: 2,372 words); (iii) between 83 and 2,324 words in Mult-Ed-EN (standard deviation: 210 words); and (iv) between 103 and 3,127 words in Mult-Ed-FR (standard deviation: 251 words). While this certainly affects the internal homogeneity of each subcorpus, the fact that the standard deviations in the English and the French subparts of each corpus are similar can probably reassure as to the cross-linguistic comparability of the texts with respect to size.

Table 5. (*Continued*)

		ORIGINAL ENGLISH	ORIGINAL FRENCH
MULT-ED	REGISTER	Quality newspaper editorials (e.g. *The Guardian, The Independent, The Times*)	Quality newspaper editorials (e.g. *Le Figaro, Le Monde, Libération*)
	TOTAL # OF WORDS	2,006,740	2,096,175
	TOTAL # OF TEXTS	4,180	4,403
	AVERAGE TEXT LENGTH	481 words	476 words
	PUBLICATION DATE*	2002-2005; 2013-2014	2002-2007; 2010-2015
TOTAL # OF WORDS (OVERALL)		4,039,846 words	4,121,547 words

* A first million words per language was compiled in the mid-2000s. At the beginning of this research project, I augmented the corpus, and thus compiled another million words per language. I decided to collect the most recently-published articles at my disposal, which is why there is a time gap between the first and the second subparts of the Mult-Ed corpus.

Despite this general comparability between the English and the French data, however, it is worth saying a quick word about the actual comparability of similar registers across languages. As pointed out by Neumann (2014: 42), "seemingly comparable registers may still display differences in their actual realization, thus […] distorting the cross-linguistic comparison". In other words, just because two texts are called 'editorials' or 'research articles' in English and French does not automatically mean that they follow the exact same register conventions (see also Mason 2001: 68 on this). Comparing English and French research articles in medicine, for example, Régent (1980; 1992; 1994) highlighted a number of cultural differences in the ways in which French and English researchers tend to present the scientific information (e.g. with respect to the ways of grouping information into paragraphs, of interacting with the other researchers, or in terms of the respective weight of information and argumentation in the development of the texts, etc.).[7] Unfortunately, according to Neumann (2014: 42):

> [T]here is no optimal solution to the question of cross-linguistic comparability of registers. If one does not want to stop attempting to make some claims about contrasted registers, this means that any claims made on the basis of a comparison of potentially incomparable registers need to be tentative.

---

7. Note, however, that Fløttum et al. (2006: 53–55) later observed the opposite tendency, since they uncovered great cross-linguistic stability in the writing conventions observed by French, English and Norwegian researchers within the disciplines of business, linguistics and medicine.

In other words, the fact that similar registers may nevertheless display more or less subtle differences is something that contrastive linguists have to live with if they want to be able to keep contrasting languages using comparable original texts. That being said, a description of the editorial and the academic registers in English and French along the three parameters of field, tenor and mode (viz. the three dimensions used by Systemic Functional linguists to describe a given register; see Chapter 4) proves reassuring with respect to the degree of comparability of the texts included in the present study.

In terms of field, defined as the subject-matter being discussed in a given situational context, English as well as French editorials are concerned with the description and discussion of a range of topical issues relating to society, politics and/ or economy. English and French quality newspaper editorials are also equivalent as regards tenor, viz. the social features of the interactants, and the relationship between them. In both cases, the author of the editorials is a senior staff member of the newspaper, whose voice is in fact supposed to reflect the stance of the newspaper as a whole towards the news event(s) being discussed (see e.g. Caffarel-Cayron & Rechniewski 2014: 20). Quality newspaper editorials stand towards the more formal end of the formality cline (González Rodríguez 2007: 57–58; Granger 2014: 62), and target a wide, public audience (Alonso Belmonte 2007: 1886; Dafouz-Milne 2008: 96). Finally, the language mode (which encompasses the channel of communication, the degree of interactivity and the rhetorical function of a text) is also comparable in the English and the French editorials: both types of texts are written, fully monologic, and allow for a relatively high degree of preparation and editing. In terms of rhetorical function, editorials in the two languages pursue a combined informational and persuasive communicative goal (see e.g. Alonso Belmonte 2007: 1886; González Rodríguez 2007: 51). On the one hand, as argued by Love (2004: 441), editorials are partly informational, in that they "recapitulate, recontextualise and connect previously reported news stories" (see also Bolívar 2002 on this). As such, they stand much closer to the 'informational' than the 'involved' end of Biber's (1988: 128) first register dimension, which measures the degree of informational focus associated with each of the twenty-two registers that he compares in his large-scale corpus study. In addition, and more importantly, newspaper editorials are also characterised by a strong argumentative component: Biber (ibid.: 148), for instance, describes editorials as an "opiniated genre intended to persuade the reader". Likewise, for Caffarel-Cayron and Rechniewski (2014: 18), editorials are a register "whose principal function is opinion formation and persuasion by argument" (see also Love 2004; Ansary & Babaii 2005; Virtanen 2005; Alonso Belmonte 2007; González Rodríguez 2007 on the persuasive functions of editorials).

In the same way as English and French editorials, English and French research articles share most of their central register features. The field of the articles is quite

naturally determined by the domain or discipline from which they were collected (Banks 2002: 2). As was already pointed out earlier, in the LOCRA corpus, cross-linguistic comparability in terms of field was ensured by including equivalent amounts of data from the same five disciplines in each language, viz. anthropology, education, political science, psychology and sociology – with each discipline representing about a fifth of the total number of words in each language. With respect to tenor, the research articles in the two languages were written by experts in the field, for an equally expert or specialised readership (Gray 2015: 10) – as opposed to the editorials, which targeted a much wider, public audience. Finally, as regards mode, research articles are monologic written texts, which allow for a high degree of careful planning and editing. Similarly to editorials, the register is characterised by a mixed informational and argumentative communicative purpose, in both French and English (see e.g. Régent 1992: 67; 1994: 3; Fløttum et al. 2006). Hyland (1998: 439–440), for instance, describes the communicative goal of research articles as follows:

> The academic writers seek to produce texts that realise specific responses in an active audience, both informing and persuading readers of the truth of their statements by seeking to "weave discourse into fabrics that others perceive as true" (Harris, 1991: 289). In other words, successful academic prose involves relating illocutionary acts to perlocutionary effects. A writer wants a message to be understood (an illocutionary effect or uptake) and to be accepted (a hoped for perlocutionary effect).

Likewise, in their contrastive study of academic discourse in English, French and Norwegian, Fløttum et al. (2006: 8) explain that "research articles are texts that primarily report and discuss empirical (including experimental) or theoretical research that has been done by the authors themselves" – thus being largely informative. However, they also stress that "[t]he final rhetorical aim of a research article is to create effects which convince the audience to such a degree that the article becomes an integrated part of a particular field's literature" (ibid.: 30; see also Lenker 2010: 42).

In summary, the description of the academic and the editorial registers along the three SFL dimensions of field, tenor and mode demonstrates that, although they may show subtle differences in the ways in which they are realised in the two languages, the two registers are in fact highly comparable with respect to their most general features. It is also important to stress that, while the academic and the editorial registers share a number of features (i.e. they are both formal written registers sharing a combined argumentative and communicative purpose), they also display a number of differences, which make their comparison relevant and interesting. For example, the respective weight of the informational and persuasive

purposes differs between academic writing and newspaper editorials. In the editorials, the informational component is somewhat secondary with respect to the persuasive one. As stated by Love (2004), newspaper editorials do not *primarily* aim to inform readers on a given event or situation. Rather, they typically deal with recently-reported news stories, with which the reader is supposed to already be familiar, and which they "analyse and/or evaluate […] to influence readers' perception of them" (ibid.: 441). In other words, while newspaper editorials *do* partly convey informational content – mostly by recapitulating events that were previously reported in more 'regular' news reports – their main communicative purpose is persuasive: they mostly intend to convince the reader to agree with the commentary that they are making on these events (ibid.: 443; see also Ansary & Babaii 2005: 278). In the research articles, by contrast, the informational component is extremely salient, as attested by the very high frequency of linguistic features that testify to the high informational density of a given text (e.g. frequent use of nouns and prepositional phrases; high type-token ratio, etc.; see Biber 1988: 128). In fact, Biber (ibid.) established that academic prose displays a significantly higher frequency of linguistic features associated with a highly informational focus than newspaper editorials. Conversely, Biber (ibid.: 148–151) found that, as regards the use of linguistic features reflecting the overt expression of persuasion, the editorials appeared to surpass academic prose. In summary, even though academic research and newspaper editorials share a combined informational and persuasive communicative purpose, research articles are more closely associated with information than the editorials, which are more strongly argumentative than academic prose. In addition to this general difference, the registers also display a range of other, more subtle divergences. These are discussed in greater detail in the analytical part of this book (Chapters 7 and 8), whenever they prove relevant for the interpretation of the corpus results.

5.2   Methodology

This section outlines the main methodological decisions taken in the framework of this research. Section 5.2.1 describes the four methodological steps that were taken to extract the English and French conjunctive markers of contrast from the comparable corpus. Section 5.2.2 then presents the statistical methods used to test the significance of the corpus results presented in Chapters 7 and 8.

5.2.1   Four main methodological steps

The extraction of English and French conjunctive markers of contrast from the comparable corpus followed a semi-automatic procedure in four main steps, listed

in Figure 9. The first stage consisted in compiling a maximally comparable inventory of conjunctive markers of contrast in the two languages, whose occurrences could then – i.e. as a second step – be extracted automatically from the corpus.
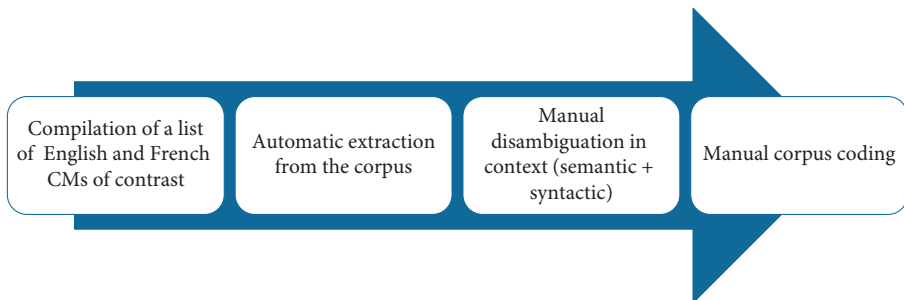
Figure 9. Main methodological steps followed to extract the data set from the comparable corpus

However, because conjunctive markers are well-known for their polysemy and polyfunctionality (e.g. the markers *still* and *yet* may be used either as concessive CMs, or as time adverbials), the fully form-based automatic extraction of the CMs in the list from the corpus was far from sufficient to ensure the reliability of the data set. Rather, the identification of English and French CMs of contrast in the corpus necessitated a very long, painstaking process of manual disambiguation in context aimed at isolating the occurrences which corresponded to the definition of conjunctive markers of contrast established in Chapter 2. More specifically, the data had to be disambiguated both (i) semantically and (ii) syntactically. On the one hand, I had to make sure to only keep the tokens which conveyed a relationship of contrast between the two discourse units – thus discarding temporal uses of *still* and *yet*, for instance, as in Examples (1) and (2). On the other hand, as explained in Chapter 2, one of the criteria that a given marker has to meet to perform a conjunctive function is that it must relate clauses or larger units. Therefore, markers expressing a relationship of contrast between two phrases, as in (3) and (4), do not qualify as markers of cohesive conjunction, and had to be weeded out as well.

(1)   Skilled in war but also in the arts of peace, he was a seaman, naval officer, farmer and administrator whose vision of what Australia could become is **still** an inspiration. (Mult-Ed-EN – The Guardian)

(2)   The WCU Project, which was introduced in 2008, aims to recruit internationally well-known faculty, with whom Korean HEIs might develop world-class academic departments and, eventually, world-class universities. Due to its short history, it is difficult to examine its outcomes **yet**. (LOCRA-EN – Education)

(3)   A panel of medical experts in Geneva concluded that, with victims facing a 70 or 80 per cent mortality rate, the Hippocratic oath to "do no harm" was insufficient reason to prevent desperate men and women from taking advantage of [risky] **but** [potentially life-saving] medicines.

(Mult-Ed-EN – The Independent)

(4)   On peut, à l'instar de Schneidermann, déplorer la surexposition médiatique du ministre [martial] <u>et</u> **cependant** [sumophobe].

(Mult-Ed-FR – Libération)

When the disambiguation process was completed, i.e. when the entire set of English and French CMs of contrast in the corpus had been identified, it was finally possible to code the data along a series of dimensions in accordance with the objectives of the corpus analyses (e.g. CM type, register, position, etc.).

In this section, focus is laid on the first two methodological steps displayed in Figure 9, viz. the compilation of a list of CMs, and the automatic extraction of the items in that list from the corpus. While the disambiguation stage may appear rather straightforward at first sight, it raised many important theoretical and methodological issues, which are described in a separate chapter (Chapter 6). Chapter 6 also presents the coding scheme devised for analysing the syntactic features of the clauses in which conjunctive markers of contrast are involved. Finally, for the sake of clarity, questions pertaining to the coding of conjunctive adjunct placement are described at the beginning of the relevant analytical chapter (Chapter 8).

5.2.1.1   *Compilation of a list of English and French conjunctive markers of contrast*

As was made clear in Chapter 2, conjunctive markers cannot be defined as a formal construct, but rather constitute a functional category made up of a diversity of words and phrases that share the function of expressing an explicit logico-semantic relationship between two clauses or larger units. The heterogeneous nature of the CM category has important implications for the corpus analyst: unlike some other linguistic phenomena, the entire set of CMs cannot be retrieved automatically from the corpus using techniques such as part-of-speech tagging, for instance (see also Castagnoli 2009: 65 on this). Not only do the different members of the category of conjunctive markers belong to a set of different grammatical classes (viz. coordinators, subordinators, adverbs, prepositional phrases) but, more importantly, far from all the members of these grammatical classes perform the function of conjunctive markers. Hence, retrieving all the coordinators, subordinators, adverbs and prepositional phrases from a part-of-speech tagged corpus would cause an unmanageable amount of noise, as only a small proportion of the automatically-extracted data would in fact prove relevant for the analysis. In view of this, the only possible solution to automatise the corpus analysis as much as

possible (that is, rather than 'read' the corpus to identify the relevant items fully manually) was to compile a list of English and French conjunctive markers of contrast to be queried automatically in the corpus.

Both the English and the French literature offer a number of ready-made inventories of CMs. What is striking, however, is the lack of comparability of the lists available. Comparing the lists of English CMs of contrast provided by Halliday and Hasan (1976: 242–243) with that put forward by Castagnoli (2009: 68–69), for example, it appears that the number of items listed in the latter inventory is about twice as high (about 60 markers) as in the former (less than 30 CMs). Such blatant discrepancies can undoubtedly be ascribed to the diversity of definitions of both conjunctive markers and contrast which are adopted in the literature (see Chapter 2 on this). In this particular case, for instance, the gap in number of CMs is partly due to the fact that Halliday and Hasan (1976) adopt a narrow view of conjunctive markers – only considering coordinators and conjunctive adjuncts as signals of cohesion – whereas Castagnoli (2009) takes a broader view, also including subordinators of contrast in her inventory. In the present study, the lists of French and English CMs were compiled by pooling some of the main lists of conjunctive markers of contrast available in the reference literature. More specifically, the English list was compiled on the basis of the following inventories of CMs: (i) Halliday and Hasan (1976: 242–243); (ii) Quirk et al. (1985: 635–636); (iii) Knott (1996: 161–169); (iv) Rudolph (1996: 4–5); (v) Chalker (1996: 26–34); (vi) Prasad et al. (2008: 65–70); and (vii) Castagnoli (2009: 68–69) – with the last of these lists being particularly helpful, as it was compiled in an English-French contrastive perspective. For French, I used the following lists: (i) Ruquet et al. (1991: 120–145); (ii) Grieve (1996); Castagnoli (2009: 68–69); and (iv) Roze et al. (2012). As the reader will have noticed, the set of resources was more limited in French than in English, but this was compensated for by the fact that French has recently benefited from a very large-scale, partly corpus-based initiative aiming to compile a maximally-comprehensive list of conjunctive markers, viz. the Lex-Conn project (cf. Roze et al. 2012).[8] Therefore, despite the more limited number of resources used to compile a list of French CMs, the availability of the LexConn database constituted a very solid – and indeed nearly sufficient – first basis for the identification of a large set of French CMs of contrast. It must also be stressed that the inventories listed here are not equivalent with respect to the criteria which underlay their compilation. Whereas some lists appear to be mainly intuition-based (e.g. Halliday & Hasan 1976; Quirk et al. 1985; Ruquet et al. 1991), others

---

8.   The LexConn inventory of CMs is available at: <http://www.linguist.univ-paris-diderot. fr/~croze/D/Lexconn.xml> (12 April 2020).

were – at least partly – corpus-driven (e.g. Knott 1996; Roze et al. 2012), while yet others had themselves been compiled by pooling previously-available lists from the literature (e.g. Castagnoli 2009).

The pooling process gave rise to a *very* mixed bag of about 100 expressions of contrast per language, some of which were found to be common across all the lists (e.g. *but, however; toutefois, en revanche*), whilst others only occurred in one or two of the lists (e.g. *contrariwise, whilst, even so; a contrario, en même temps, plutôt*). Faced with this very wide variety of expressions, I then had to go through a careful selection of the items which in fact corresponded to the definition of conjunctive markers of contrast adopted in this study. Two main types of filters were applied to these fairly heterogeneous lists, viz. a semantic and a grammatical one. From a semantic point of view, I only kept the markers which corresponded to the definition of contrast established in Chapter 2. In other words, only the CMs that may express a meaning of opposition (e.g. *by contrast; à l'inverse*), concession (*nevertheless, toutefois*) or correction (*instead, plutôt*) were included in the final list. By contrast, markers of reformulation (e.g. *better, more accurately*), exception (e.g. *only, si ce n'est que, à ceci près que*) or dismissal (e.g. *in any case, at any rate*) were excluded. In addition, the markers which may, in some very specific contexts, express contrast, but are used to express (an)other meaning relation(s) in a very large majority of the cases – such as *and* in (5), which may easily be replaced by *yet*; or *quand* in Example (6) – were also excluded from the list. It was deemed that the proportion of uses of these markers that would in fact express a relation of contrast was unlikely to be worth the work required to manually disambiguate such instances in context. This is especially true as, in many such examples (including (6) below), the relation of contrast is marked a second time by a specifically contrastive marker (viz. *au contraire*).

(5)   Provided basic safety standards are set, and children are protected, the only protections that adults require are information, information and more information. The EU should legislate along these lines, but only if it can show that existing law in its member states is inadequate – **and** that is as unproven as the beneficial properties of Rhodiola rosan.

<div align="right">(Mult-Ed-EN – The Independent)</div>

(6)   Le premier ministre dit qu'il se passerait volontiers de cette référence **quand** la maire de Lille y est, au contraire, attachée corps et âme.

<div align="right">(Mult-Ed-FR – Le Figaro)</div>

Finally, I also excluded the markers which were not strictly conjunctive, but whose meanings (also) exhibited a certain 'disjunctive flavour', in the sense that they provided some information on the stance of the writer towards the message. Accordingly, markers such as *of course, in fact, clearly; en réalité, certes, évidemment,*

malheureusement were not kept for the corpus analysis, even though they may admittedly be used to signal a contrast between two discourse units, as in (7) and (8) (see e.g. Thompson & Zhou 2001; Degaetano-Ortlieb 2015 on the conjunctive function of disjuncts). The main rationale for this decision is that, among these disjunctive devices, it can be very difficult to draw the line between the items that may and may not be considered to fulfil a truly conjunctive function (compare for example: *clearly* vs *it is clear/true that*). In addition, it would also have been very hard to collect *all* the disjunctive adverbs that may be used to express contrast, and thus achieve a comparable list of these items across languages. A final argument justifying this decision is that in many cases, these items are not the ones which convey the main relation of contrast: rather, they introduce and/or reinforce the first part of a concession, while the real concessive relation is conveyed by another marker, as in (9).

(7) For the Doha round to succeed, this mercantilist mindset needs to change, and quickly. Although there is no formal date for concluding the negotiations, 2006 is a widely agreed deadline since George Bush's fast-track negotiating authority runs out in 2007. Trade ministers will therefore need to have at least a skeleton agreement in agriculture, industrial goods and services by December 2005 when they gather in Hong Kong for the World Trade Organisation's biannual meeting. **Unfortunately**, there are several reasons to fear that the Doha negotiators will not get there.

<div align="right">(Mult-Ed-EN – The Economist)</div>

(8) Certains diront que ce régime est attentatoire aux libertés publiques. Mais la sécurité des innocents n'est-elle pas la première des libertés ? Quand des fanatiques ne pensent qu'à nous anéantir pour ce que nous sommes - des Occidentaux -, nécessité fait loi. **Malheureusement**, la réforme pénale de Mme Taubira contre la récidive, examinée ces prochains jours, n'obéit pas à cette impérieuse nécessité.   (Mult-Ed-FR – Le Figaro)

(9) En collant au chef de l'État, Hollande a **certes** su éviter la faute, <u>mais</u> il n'a pas échappé au rôle peu enviable de commandeur mutique face au président en action.   (Mult-Ed-FR – Le Figaro)

From a grammatical point of view, the list of CMs used in the present study was limited to the three categories of coordinators (*but, mais*), subordinators (e.g. *while, alors que*) and conjunctive adjuncts (e.g. *nevertheless, on the other hand; toutefois, au contraire*). In other words, I did not include more peripheral elements such as contrastive verb phrases (e.g. *il n'en reste pas moins que; it remains that*, etc.) or anaphoric sequences (e.g. *in spite of this; malgré cela*, etc.), sometimes referred to in the literature as 'alternative lexicalizations' of discourse relations (Prasad et al. 2010) or 'secondary connectives' (Rysová & Rysová 2014). As explained by

Prasad et al. (2010: 1024), such elements are part of a "lexically open-ended class of elements which may or may not belong to well-defined syntactic classes". As was the case for disjuncts, the open-ended character of the category of alternative lexicalizations would have made it very difficult to know when to stop in the compilation of an inventory of such forms. Importantly, the near impossibility to reach comprehensiveness with respect to these items may in turn have impeded the cross-linguistic comparability of the English and French lists of CMs. As a result, it was felt preferable to exclude such markers altogether. This decision was supported by quantitative criteria, as such elements have been demonstrated to account for a very small proportion of the explicit markers of logico-semantic relations in authentic data. Based on the manual annotation of about 50,000 sentences from original newspaper texts in Czech, for example, Rysová and Rysová (2014: 456) found that such realisations only account for about 5% of the total number of explicit conjunctive markers in their corpus, with the remaining 95% of the relations being expressed through what they call 'primary connectives', viz. conjunctions and adverbs (see also Dupont 2013: 86 for similar results on English and French). In addition, it is also worth noting that the decision to exclude anaphoric expressions of logico-semantic relations between discourse units – such as *in spite of this* or *malgré cela* – is in accordance with Danlos et al.'s (2015b: 7) view that CMs which contain an anaphoric element should be "considered as non-connective", in view of the fact that their meaning is partly compositional.

At the end of the process of selection of CMs along both semantic and grammatical criteria, the final list of conjunctive markers of contrast consisted of 42 English and 50 French items. The list is provided in Table 6, where the CMs are ordered alphabetically within each category.

Table 6. List of English and French conjunctive markers of contrast to be queried in the corpus

CATEGORY	ENGLISH	FRENCH
COORDINATORS	But	Mais
SUBORDINATORS	Albeit	Alors même qu(e)
	Although	Alors qu(e)
	Even if	Au lieu qu(e)
	Even though	Bien qu(e)
	For all that	Cependant qu(e)
	Though	En dépit qu(e)
	Whereas	En même temps qu(e)
	While	Malgré qu(e)
	Whilst	Même s(i)
		Nonobstant qu(e)

Table 6. (*Continued*)

Category	English	French
		Quand bien même qu(e)
		Quand bien même
		Quoiqu(e)
		Si/s'
		Tandis qu(e)
		Tout en
Conjunctive adjuncts	All the same	A contrario
	Anyhow	À la place
	Anyway	À l'inverse
	Anyways	À l'opposé
	At the same time	À titre de comparaison
	By comparison	Au contraire
	By contrast	Au lieu
	By way of comparison	Bien au contraire
	By way of contrast	Cependant
	Contrariwise	Comparativement
	Conversely	D'autre part
	Even so	De l'autre
	However	De toute façon
	In comparison	De toute manière
	In contrast	De toutes (les) façons
	Instead	De toutes (les) manières
	Just the same	D'un autre côté
	Meanwhile	En même temps
	Nevertheless	En revanche
	Nonetheless/none the less	En comparaison
	Notwithstanding	Inversement
	On the contrary	Malgré tout
	On the other	Néanmoins
	On the other hand	Nonobstant
	Oppositely	Or
	Quite the contrary	Par comparaison
	Quite the opposite	Par contre
	Rather	Plutôt
	Regardless	Pourtant
	Still	Quand bien même
	Though	Quand même
	Yet	Tout de même
		Toutefois

5.2.1.2 *Automatic extraction of the conjunctive markers from the corpus*

Once an inventory of conjunctive markers of contrast had been compiled in both languages, I was able to extract all the occurrences of the markers in this list

automatically from the corpus. This was done using WordSmith Tools's *Concord* concordance tool (Scott 2012), which makes it possible to retrieve all the occurrences of a (set of) forms from one or several corpora and visualise them in context. In line with the objectives of the present study, all three categories of conjunctive markers were extracted from the Mult-Ed subcorpus of editorials, whereas for reasons of feasibility, the LOCRA subcorpus was only used to investigate conjunctive adjuncts of contrast. Figures 10 and 11 below illustrate the type of output obtained using Concord's fully form-based method of extraction. As appears clearly from these screenshots, the automatic extraction of the occurrences only constitutes a minor part of the analysis and must be followed by a long process of manual disambiguation in context. For example, some of the concordance lines in Figure 10 had to be weeded out manually, as they correspond to temporal or intensifying uses of the adverb *yet*. Likewise, all the nominal uses of *or* (i.e. referring to the French noun *gold*) in Figure 11 had to be discarded. The manual disambiguation of the data is the object of Chapter 6.

. In those three years rent rises would be limited by a formula, as yet unspecified, related to inflation, earnings or average market , was not worthy of negative comment. That can never be true. Yet it can be a convenient concept for those individuals who are . So did the occasionally eccentric actions of judges. And yet crime, generally, is falling, and the prisons are full. Prison, if it replacing it with a "British Bill of Rights", the contents of which are yet to be confirmed. The best way of managing each of those spending cuts, or about £17bn, has shocked French people, and yet it is small beer compared with what George Osborne has in the festive season thanks to pre-Christmas downpours, and there is yet more to come. The Environment Agency issued no fewer than the Cobra emergency committee to discuss a national response to yet more flooding. This time, it is the South and West coasts of demands mounting to a negation of democracy and are a recipe for yet more strife. It is time for both sides to step back from the human ingenuity, with our ability to make plans and write poetry. Yet the most complex structure in the known universe – as it is makes unravelling the genome, for example, look like child's play. Yet the scientific quest should be applauded and encouraged, new-found belief in market-oriented "social democracy". And yet Mr Hollande agonised over the choice. The problem, as he with which the crime of sexual assault needs to be taken. Yet at the same time less than 20 per cent of rapes are reported,

Figure 10. Extract of the concordance of English *yet* in Mult-Ed (WST6)

de guérison particulière. En Inde, on attribue des vertus prophylactiques à l'or, symbole du soleil et du Gange, et l'ambre est dit guérir la jaunisse consiste en une relation subjective, quasi émotionnelle en quelque sorte. Or, avec d'autres analystes de la parenté, nous nous posons la question l'ordre dans le grand chaos que ces parentés non régulées ont engendré. Or, s'il y a eu un regain d'intérêt, en France notamment, pour tenter de leurs approches théoriques de distinguer les relations sexuelles du mariage. Or, comme le note Jeanette Edwards (2004) ou Corinne Fortier (2005), significative au niveau moral, entre ces types de dons intrafamiliaux. Or, l'écart relevé par Thompson s'observe dans plusieurs pays, dont le de gamètes et autorise les dons directs entre donneurs proches et connus. Or, il faut savoir que les dons de sperme de tous donneurs doivent être » actuelle de la parenté particulièrement renforcée en Occident. Or, on constate, au contraire, avec la révélation de ces façons de faire bords, ces derniers apparaissant de façon significative sur les deux bords. Or pour les assemblages lithiques du niveau I de Cagny-l'Epinette, il a été coléoptères, orthoptères et arachnides représentés dans les poids à peser l'or ashanti et baoulé, en particulier les « insectes moulés de la collection des insectes devenus des documents ethnographiques, les poids à peser l'or apparaissent donc comme des objets ethnographiques ou des œuvres ambitionne de publier un ouvrage intitulé « Bêtes et plantes des Dogon ». Or, après la guerre, ses recherches sont reprises et poursuivies, par sur les formes d'homosexualité institutionnalisée observées dans les mines d'or sud-africaine (Moodie et al. 1988 ; Harries 1990), pour s'être focalisés sociale largement délaissés par la suite, dont bien sûr celui de la sexualité ; de telles approches sont aujourd'hui frappées de disgrâce. Pourtant, l'un

Figure 11. Extract of the concordance of French *or* in LOCRA (WST6)

Once the data had been extracted automatically from the corpus, it was exported in Excel (.xlsx) format to allow for as much flexibility as possible in the ensuing disambiguation and coding processes. While WordSmith Tools's (henceforth: WST) concordancer is very user-friendly, it only allows for one coding layer per concordance line. This was far from enough for the type of analyses performed in the present study, where each concordance line is coded along a series of variables, with a view to uncovering possible interactions between these variables on the use of conjunctive markers of contrast (see Chapters 7 and 8 for more details).

Other corpus coding software was also tested to assist me in the coding process, but none of them offered the same amount of flexibility as the combination of WST's Concord tool with Excel. For instance, I attempted to use the corpus coding tool designed specifically for Systemic Functional corpus analyses, viz. the UAM Corpus Tool (2009). While this tool allowed for very efficient multilayer coding procedures, it did not allow for the combined automatic extraction *and* coding of specific linguistic items in context. Thus, for example, it was possible to (i) extract all the instances of *however* from the corpus and display them in a concordance; (ii) code and annotate each text and its subparts separately along a series of dimensions according to a multilayer coding system of your own choice; but it was not possible to directly code the automatically-extracted instances of *however* in the tool. In other words, in order to code my data by means of the UAM Corpus Tool, I would have had to read all the texts and manually retrieve all the CMs in my corpus, which was not feasible in view of the amount of data treated in the present study. Finally, it must be noted that, due to their very high frequency in the corpus, the English coordinator *but* (over 14,000 hits in Mult-Ed), the French coordinator *mais* (nearly 9,000 hits in Mult-Ed), and the French subordinator *si* (c. 6,000 hits in Mult-Ed) were analysed using random samples of 2,000 occurrences per marker (computed via the *delete > reduce to N* functionality in *Concord*). The sample size of 2,000 occurrences was chosen because it corresponds to the frequency of the next most common CMs in the corpus, i.e. the most frequent CMs for which full samples were analysed.

5.2.2   Statistical methods

This section presents the main statistical procedures followed in the present study. Each analytical chapter in this book uses different statistics, in accordance with the type of comparisons carried out. More specifically, as Chapter 7 is mostly restricted to monofactorial comparisons of frequencies, the statistical significance of the differences observed in that part of the analysis is assessed using chi-square tests of independence (see Section 5.2.2.1). Chapter 8, on the other hand, takes a multifactorial stance towards the comparison of English and French, attempting to predict conjunctive adjunct placement based on a series of predictors, such as register or lexis. The statistical method used for this purpose is that of Classification and Regression Trees, described in Section 5.2.2.2. In both cases, the statistical analyses are performed using the open-source statistical software R (R Development Core Team 2008).[9]

5.2.2.1 *Frequency comparisons: The chi-square test of independence*

5.2.2.1.1 Taking the internal variability of the corpus data into account

As stated above, the type of research questions asked in Chapter 7 mostly pertain to monofactorial comparisons of frequencies of the kind: "what language displays the higher frequency of explicit conjunctive markers of contrast?" Several types of statistical tests make it possible to determine the degree of significance of a given frequency difference observed in corpus data. Ideally, researchers in corpus linguistics ought to choose a statistical test which takes account of the internal variation characterising most corpora. In recent years, corpus linguistics has witnessed a growing awareness of the fact that "corpora are inherently variable internally" (Gries 2006: 110), in the sense that the range of writers and/or speakers having authored the various texts in a corpus are likely to display partly idiosyncratic linguistic behaviours. Accordingly, comparing corpus frequencies across two or more corpora taken as wholes can be misleading, as it runs the risk of disguising "the fact that most phenomena of interest will yield different results when investigated in different parts of a corpus" (ibid.). In corpus linguistics, two main types of test tend to be used to compare corpus frequencies while taking account of the internal variability of the data. The first one is the t-test, which can be used "to test whether two group means are different" (Field et al. 2012: 368). The problem with the t-test is that it requires the corpus data to be normally distributed, viz. distributed "symmetrically around the centre of all scores [so that] if we drew a vertical line through the centre of the distribution then it should look the same on both sides [and take the form of a] bell-shaped curve" (Field et al. 2012: 19). Yet, as revealed by a Shapiro-Wilk test – which is one of the tests that can be used to test the degree of normality of the distribution of corpus data; see Gries 2013: 162–163) – the data in the Mult-Ed corpus (viz. the corpus that is at the centre of most analyses in Chapter 7) is very far from normally distributed, which rules out the use of a t-test for assessing the degree of statistical significance of the differences observed in this corpus.

Rather, the Mult-Ed corpus thus requires the use of a non-parametric test, viz. a test which does not assume the normal distribution of the data. The non-parametric equivalent to the t-test is known as the Wilcoxon rank-sum test (see Field et al. 2012: 655–673 for a description). The problem is that, even though the Wilcoxon rank-sum test does not require the data in the corpus to be normally distributed, there are still restrictions to its applicability. In particular, the accuracy of the Wilcoxon rank-sum test has been demonstrated to be strongly influenced by the size of the sample analysed: as pointed out by Zimmerman (2003), the tendency of the Wilcoxon test to report a significant effect when there is none (a phenomenon known as the "Type I error") tends to increase considerably with the size of the samples being compared. This is an effect that Zimmerman observes

when comparing the robustness of the Wilcoxon test on samples ranging from 20 to 200 observations. In comparison, the Mult-Ed corpus displays over 8,000 observations, which seriously calls into question the applicability of the Wilcoxon rank-sum test on this particular data set. Because the data in Mult-Ed was not distributed normally, and because the number of observations in the corpus was too large to be appropriately handled by the Wilcoxon rank-sum test, it was unfortunately not possible to use a statistical test which took account of the variation in CM usage across the texts in the corpus. Rather, the most suitable option with respect to the properties of my data set was to use chi-square tests of independence, which treat the corpora compared as aggregate data.[10] Compared to the t-test and the Wilcoxon rank-sum test, the chi-square test is therefore less precise in the differences that it uncovers. However, in view of the features of the corpus data used here, it also appears to be a more accurate alternative to the other two tests – a quality which, in my opinion, should take precedence over preciseness.[11]

5.2.2.1.2 Description of the chi-square test of independence

The Pearson chi-square test of independence is an inferential statistical test which assesses whether a given frequency distribution observed in (corpus) data differs significantly from the distribution that you would have observed if the frequencies had simply been determined by chance. More precisely, as explained by McEnery et al. (2006: 55):

> The chi-square test compares the difference between the observed values (e.g. the actual frequencies extracted from corpora) and the expected values (e.g. the frequencies that one would expect if no factor other than chance were affecting the frequencies […]). The greater the difference (absolute value) between the observed values and the expected values, the less likely it is that the difference is due to chance. Conversely, the closer the observed values are to the expected values, the more likely it is that the difference has arisen by chance.

---

10.   Note that another alternative may have been to use Log-likelihood tests. As explained by Bestgen (2017: 35–36), the chi-square and the Log-likelihood tests work in very similar ways and are based on an identical (viz. chi-square) distribution. As a result, they frequently return highly comparable results. In fact, for information purposes, all the chi-square comparisons carried out in the present study were also tested by means of the Log-likelihood test. For each of these comparisons, the two tests returned the exact same values. Therefore, it seems as though in the present study, both tests may have been used interchangeably.

11.   Note, however, that the data set used in the present study violates the independence assumption of the chi-square test, since some of the texts contain more than one instance of conjunctive markers. Although ignoring this violation is widespread in corpus linguistics, this is an issue that is worthy of mention.

This section will not expand on the mathematical principles behind the chi-square test, as this test may be – and indeed has been – computed automatically in R (using the 'chisq.test' function in the 'stats' package). Rather, it provides the necessary information to be able to interpret the chi-square values provided in Chapter 7.

First of all, in order to understand the mechanisms underlying the chi-square test, it is useful to specify that this test is computed on the basis of a contingency table, which provides a cross-classification of the data across the relevant categories or levels of the variables investigated. In other words, in a contingency table "[t]he levels of each variable are arranged in a grid, and the number of observations falling into each category is noted in the cells of the table" (Field et al. 2012: 915). By default, the output of a chi-square test provides three main pieces of information, viz. (i) a chi-square value; (ii) the number of degrees of freedom associated with the test; and (iii) a p-value, specifying the degree of statistical significance of the difference(s) assessed by the test. The degrees of freedom associated with a given chi-square test are directly dependent on the number of cells in the contingency table, and correspond to: [the number of rows in the table – 1]*[the numbers of columns in the table – 1]. This information is the one that makes it possible to interpret the chi-square value returned by the test, as it determines the threshold of significance of a given chi-square value. The more degrees of freedom (i.e. the higher the number of cells in the contingency table), the higher a given chi-square value needs to be to reach statistical significance. For example, a chi-square test with only one degree of freedom only requires a chi-square value of 3.84 to reach statistical significance at a $p \leq 0.05$ level. With two degrees of freedom, on the other hand, the chi-square value needs to be equal to or larger than 5.99 to be significant at a 0.05 level, against 7.82 with three degrees of freedom, and so on. Fortunately, this is a conversion that one does not usually have to make, as automatically-computed chi-square tests directly provide us with a p-value indicating the degree of statistical significance of the difference observed in the corpus data. In the present study – and in accordance with the convention usually followed in the Humanities – the significance threshold is set at $p \leq 0.05$, which means that a given difference is considered statistically significant when there is only a 5% (or lower) probability that it is due to chance (Field et al. 2012: 51).

In addition to the automatically-computed chi-square value, the number of degrees of freedom and the p-value that are provided by chi-square functions, it is also useful to provide an effect size for each chi-square test, in order to assess how strong the effect of the variable investigated is on the observed results. As stated by Field et al. (2012: 57), "just because a test statistic is significant doesn't mean that the effect it measures is meaningful or important". The main problem with chi-square and p-values is that they are very sensitive to the size of the sample analysed (Gries 2013: 185). This entails that the size of these values is not a reliable measure to assess the strength of a given effect on the results, as they tend

to increase with the size of the sample (see also Levshina 2015: 129–130; Plonsky & Oswald 2014: 879 on this). Effect sizes, by contrast, are a "standardized measure of the magnitude of [an] observed effect" (Field et al. 2012: 57). As a standardised measure, and unlike chi-square and p-values, effect sizes are not affected by the size of the samples analysed, which also makes it possible to compare them across different studies using distinct data sets and/or investigating the effect of different variables (ibid.).

Effect sizes are usually quantified with a correlation coefficient – viz. φ in the present study; see Gries (2013: 185) – that may range between 0 (no effect) and 1 (very strong effect). Traditionally, effect sizes have been interpreted according to Cohen's (1988; 1992) scale, which was first put forward in the framework of behavioural science, and assumes that: (i) an effect size value of 0.10 corresponds to a small effect; (ii) a value of 0.30 corresponds a medium effect; and (iii) a value of 0.50 or more reflects a large effect of the variable investigated. Recently, however, some researchers have called into question the relevance of interpreting effect sizes according to such a 'one-size-fits-all' kind of scale. For Plonsky and Oswald (2014), for instance, effect sizes ought to be interpreted in relation to the field in which they are obtained. Taking the example of Second Language research, Plonsky and Oswald (ibid.: 889) demonstrate that "Cohen's scale underestimates the range of effects typically obtained in L2 research", and that "quantitative L2 research produces substantially larger effects than those typical in many other fields" (ibid.: 890). Accordingly, they put forward an alternative scale that proves more appropriate for the interpretation of the effect sizes typically obtained in their field of research. The problem is that, to my knowledge, no study in corpus-based contrastive linguistics has yet systematically reported the effect sizes associated with their chi-square results. Consequently, I do not have any appropriate yardstick at my disposal to interpret the effect sizes obtained here. However, in this study, I will consistently report my effect sizes, so that they can serve as a first basis for interpretation and comparison in future corpus-based contrastive research.

A final piece of information that needs to be provided when reporting the results of a (significant) chi-square test are so-called standardised residuals. Especially with contingency tables that have more than one degree of freedom (i.e. which have more than two rows and two columns), the chi-square and the p-values do not provide sufficient information to properly interpret the results of a chi-square test. Instead, the researcher needs to clearly identify which cells of the table contribute to the statistically significant result reported by the test. This kind of information is provided by the standardised residuals, which specify the contribution made by each cell of a contingency table to the overall chi-square value. The significance threshold for such standardised residuals is 1.96. If the standardised residual associated with a given cell in the contingency table has an absolute value of 1.96 or more, it means that the frequency reported in this cell diverges statisti-

cally significantly from the frequency that would have been obtained by chance (i.e. the expected frequency). The larger the standardised residual, the more significant the contribution of the cell to the overall chi-square value. If the value is smaller than 1.96, on the other hand, the frequency in the cell does not diverge significantly from the expected frequency. In addition, each standardised residual value is accompanied by either a plus or a minus sign. This information tells us about the direction of the observed effect. While a standardised residual topping +1.96 indicates that a given cell in the contingency table displays a *higher* frequency of the phenomenon under study than would be expected, a smaller residual than –1.96 tells us that the cell displays a significantly *lower* frequency than the one which may have been obtained by chance. Importantly, it must be specified that standardised residuals are only relevant when the overall chi-square value has returned a statistically significant result.

5.2.2.2 *Classification and Regression Trees (CART)*

As opposed to the analyses carried out in Chapter 7, Chapter 8 takes a more multifactorial stance towards the study of CM use in English and French. More specifically, Chapter 8 investigates the placement patterns of English and French conjunctive adjuncts of contrast (e.g. *however, nevertheless, cependant*), assessing the combined influence not only of language, but also register and lexis on the phenomenon. Such a research agenda requires the use of multifactorial statistical methods, i.e. methods which make it possible to tease out the respective influence of a range of variables on the phenomenon at hand, while also shedding light on possible interactions between them. The method used to assess the combined influence of language, register and lexis on conjunctive adjunct placement in English and French is that of Classification and Regression Trees.

Classification and Regression Trees (CARTs) are an inferential statistical method that aims to predict a given outcome (e.g. the position of conjunctive adjuncts) from a number of predictors (e.g. language, register and lexis) on the basis of a given data set (see Breiman et al. 1984 for a book-long description; see also Berk 2016: chap. 3 for a more accessible summary). The basic principle underlying the CART method is that, based on the data set that has been fed to it, it attempts to identify the best predictor of the outcome at hand from among a range of independent variables, i.e. the predictor that leads to the largest reduction of the possible outcomes (Berk 2016: 130).[12] This best predictor gives rise to a first binary distinction within the data set, which can be read as the first stage in a deci-

---

12.   Note that within the CART method, classification trees are used to predict the values of a categorical dependent variable, while regression trees are used to predict the values of a numerical dependent variable. As the position of conjunctive adjuncts is a categorical variable, the present study resorts to classification rather than regression trees.

sion tree (e.g. if the language is English, I am more likely to observe position X; if the language is French, I am more likely to observe position Y). The search for the best predictor is then repeated on each of the two subsets of the data, which themselves give rise to new splits: the process is repeated recursively, until the CART algorithm can no longer find any splits that would lead to a significantly more accurate prediction of the outcome, leading to more meaningful improvements to the predictive model (ibid.: 132). The fact that each stage of the CART process aims at identifying the single best possible predictor of a given outcome implies that the most influential factors appear at the top of the tree, while the factors inducing lower divisions have a more limited role in predicting the outcome. In other words, the earlier a given predictor gives rise to a split in the tree, the greater role it has in influencing the phenomenon under investigation.

Finally, it is important to stress that CART trees may take two different forms, viz. they may be reported either in a 'pruned' or in a 'non-pruned' version. As a first step, the CART method always draws a very detailed tree, which provides the best possible representation of the data set used to build the model. This first version corresponds to the most complete, or 'non-pruned' version of the CART tree. In a second stage, however, the method applies a process of cross-validation on the data, with a view to building a tree which would also be valid if another, comparable data set were to be analysed. This usually leads to the deletion (or pruning) of branches that were present in the initial version of the tree, but appear to be somewhat specific to the data set on the basis of which the model was built, rather than being generalisable to other, similar situations. In other words, whereas non-pruned trees are useful to provide an optimised description of the data set, pruned trees have greater predictive power than their non-pruned equivalents: they make it possible to forecast what might happen in a comparable data set to the one that has been used to build the model (see Berk 2016: 157–159 for more details on pruning and cross-validation with respect to CART trees). Given that researchers usually aim at formulating – at least partly – generalisable observations on the phenomena investigated, CART trees are typically reported in pruned form. However, it may still be informative to have a look at the non-pruned versions of the trees as well, so as to gain deeper insights into the specificities of a given data set.

5.3 Conclusion

This chapter has provided a general description of the corpus resources used in this study, as well as the methods adopted to analyse them. After demonstrating the necessity of resorting to comparable – rather than translation – data in an onomasiological contrastive study of conjunctive markers, the first part of the chapter presented the corpus used to compare CMs of contrast in English and French,

laying special emphasis on issues of cross-linguistic comparability between the texts included in the data set. The comparable corpus is made up of two subcomponents representing two distinct language registers, viz. newspaper editorials and research articles. The features of each of these registers were described along the three SFL dimensions of field, tenor and mode, which provided reassuring evidence concerning the degree of cross-linguistic comparability of these two text types.

The second part of the chapter briefly described the four methodological steps followed in this study – viz. (i) compilation of a list of CMs of contrast; (ii) automatic extraction of the CMs from the corpus; (ii) semantic and syntactic disambiguation in context; and (iv) manual coding of the data – and focused more particularly on the first two of these stages. The list of English and French CMs of contrast was compiled by pooling a number of ready-made inventories of markers available in the literature. This gave rise to a very mixed bag of English and French contrastive expressions, which was subsequently filtered according to both semantic and grammatical criteria, to only keep the markers which corresponded to the definition of conjunctive markers of contrast adopted in this book. The final inventory included 42 English and 50 French markers of contrast, whose occurrences (39,894 in total) were extracted automatically from the corpus using the WordSmith Tools 6 corpus analysis software, and exported in Excel format for subsequent manual disambiguation (see Chapter 6) and coding (see Chapters 6 and 8). The final section of the chapter described the statistical methods adopted in the corpus analyses presented in Chapters 7 and 8, namely the Pearson chi-square test of independence, and the multifactorial method of Classification and Regression Trees.

CHAPTER 6

Beyond automatic extraction

Semantic disambiguation and syntactic segmentation

As explained in the previous chapter, the fully form-based automatic extraction of (potential) markers of contrast from the corpus only constitutes one of several stages required to obtain the final set of elements truly functioning as conjunctive markers of contrast. After the markers in my list had been retrieved automatically from the corpus data, their occurrences had to be disambiguated manually along two main dimensions. Firstly, because some markers in the list of CMs compiled for the purposes of the present study are polyfunctional (in that they may express a range of other meanings in addition to contrast), the data had to be disambiguated semantically. The objective was to only keep the items which express a meaning relation of contrast – thus discarding temporal uses of *while* and *alors que*, for instance. This process is described in Section 6.1. Secondly, as already mentioned earlier, one of the defining features of conjunctive markers is that the segments that they relate should correspond to clauses or larger units, while markers linking smaller units – such as phrases – are not considered to perform a conjunctive function. As a result, the data also had to be disambiguated syntactically, in order to weed out the occurrences of markers operating below clause level. This procedure – which is usually referred to as 'discourse segmentation' – is the object of Section 6.2. It is worth recalling here that, in line with the objectives of the present study, all three categories of conjunctive markers (viz. coordinators, subordinators and CAs) are investigated in the Mult-Ed subcorpus, whereas the analyses carried out on the basis of LOCRA only pertain to conjunctive adjuncts. It should also be stressed that, while it is true that the discussions included in this chapter are somewhat fastidious, they are also extremely important: the criteria of selection presented here had a decisive impact on the data set kept for the analysis, and on the results of the corpus study. Consequently, it was crucial that these criteria be made explicit, to make clear exactly what types of units are included in the present study.

6.1   Semantic disambiguation

6.1.1   The polyfunctionality of conjunctive markers

Conjunctive markers are notoriously polyfunctional linguistic items: as underlined by Meyer et al. (2011: 194), "the presence of a connective does not unambiguously signal a specific discourse relation. In fact, many connectives can indicate several types of relations between sentences, i.e. they have several possible 'senses' in context". Many items in the list of markers compiled for the purposes of this study are not restricted to a meaning of contrast. For example, both the English marker *while* and the French CM *alors que* may be used to convey a meaning of temporality – as in (1) and (2) – as well as contrast – as in (3) and (4). As the present research only investigates CMs expressing a logico-semantic relation of *contrast* in English and French, it was necessary to manually screen all the occurrences of the markers investigated in this study to only keep the contrastive uses of these CMs.

(1)   Assad was a dictator and – more importantly – an unfriendly one. The chance to put an end to his regime appeared too good to miss. But **while** policymakers in Washington, London and Paris were busy concocting their schemes, they forgot to factor in Plan B: what if the Assad regime declines to fold, or if a democratic revolt against the regime morphs into a sectarian crusade?   (Mult-Ed-EN – The Independent)

(2)   Quinze jours après les attentats, **alors que** Madrid pleure encore ses morts, la police espagnole semble en mesure de dire qui sont les commanditaires et les auteurs du carnage du 11 mars.   (Mult-Ed-FR – Le Figaro)

(3)   The message for David Cameron is more nuanced. **While** he did badly, he did not do nearly as badly as he might have done and the opinion polls continue to show that the general election is neck and neck.
                                                                        (Mult-Ed-EN – The Telegraph)

(4)   Même si elles divergent quant au degré de l'intégration, la France et l'Allemagne préconisent une Europe politique, **alors que** la Grande-Bretagne se contenterait d'un grand marché.   (Mult-Ed-FR – Le Figaro)

There are in fact two ways in which conjunctive markers may be polyfunctional: firstly, a given marker may have the potential to express different meanings in distinct contexts. This is what we observe in Examples (1) to (4) above where, depending on the context, *while* and *alors que* signal a relation of contrast *or* temporality. This phenomenon is referred to as 'sequential multifunctionality' by Petukhova and Bunt (2009; see also Meyer et al. 2011: 195; Cartoni et al. 2013: 66). Secondly, a given CM may express several meaning relations simultaneously in a single occurrence – thus displaying what Petukhova and Bunt (2009)

call 'simultaneous multifunctionality'. This is what happens in Example (5) below, where *while* expresses both temporality and contrast at the same time: on the one hand, the writer is opposing the poverty of the population to the wealth of President Mugabe's cronies; on the other hand, s/he is stating that the situations described in S1 and S2 are happening at the same time.

(5)   Zimbabwe provides a dramatic illustration of how statist economic policies, corruptly enforced, swiftly impoverish. In the past five years, Mr Mugabe's contempt for property rights has made half the population dependent on food aid, **while** his cronies help themselves to other people's land and savings, and build helipads for their own mansions.

<div align="right">(Mult-Ed-EN – The Economist)</div>

Both types of polyfunctionality were observed in the corpus data. In this respect, it needs to be stressed that, while the polyfunctionality of CMs constitutes a considerable resource for language users, enabling them to make explicit the sometimes very complex relationships that may hold between two events or situations, it also poses major challenges for the discourse analyst as it can give rise to highly ambiguous occurrences, which are very difficult to assign to a well-defined semantic category (see among others: Miltsakaki et al. 2005; Spooren & Degand 2010; Meyer 2011; Cartoni et al. 2013 on this). Such ambiguities are in fact presented as one of the main challenges of discourse analysis by Schiffrin et al. (2015: 7), who underline the difficulty of selecting the right interpretation of the relationship between two discourse units in some contexts: "[s]ince more than one meaning can be created, how do we decide which meaning is intended, is justifiable, and/or makes the most sense?"

With respect to CMs of contrast, two main types of ambiguities were especially recurrent in the data, viz. (i) ambiguity between contrast and temporality; and (ii) ambiguity between contrast and addition. Such cases are discussed and exemplified in greater detail in Section 6.1.2. But before discussing this range of problematic cases, it must also be made clear that a fair proportion of the markers investigated here were relatively easy to deal with: CMs such as *nevertheless* or *pourtant*, for example, (nearly) systematically express a logico-semantic relation of contrast between the segments that they link. The inclusion of such instances in the data set was therefore fairly straightforward. In some cases, a single marker could express different subtypes of contrast: French *au contraire*, for example, may express either opposition – as in Example (6), or correction – as in (7). However, such cases of polyfunctionality did not pose any problems for the present study, which includes all three subcategories of contrastive meanings, while not seeking to annotate each occurrence in the corpus with respect to the subtype of contrast expressed by the marker.

(6)   Précisons d'abord que certaines de ses lettres à Else sont très pénibles à lire :
        un degré zéro d'érotisme s'y dégage d'auto-humiliations et d'exaltations
        exagérées. D'autres lettres, **au contraire**, sont très belles.

                                                            (LOCRA-FR – Sociology)

(7)   Ce document, expurgé de ses passages les plus abjects, ne vise pas à
        entretenir on ne sait quel voyeurisme malsain sur une tragédie qui a
        bouleversé la France. Il pose **au contraire** une nouvelle brique factuelle et
        dissipe une part de l'épais brouillard qui continue à entourer l'affaire.

                                                            (Mult-Ed-FR – Libération)

6.1.2   Contrast and other meaning relations: Some areas of overlap

The semantic disambiguation of the corpus data revealed that contrast is far from
a clear-cut semantic category, but instead overlaps with other meaning relations.
This section discusses and exemplifies some of the most recurrent cases of ambi-
guity encountered in the data. Probably the most common area of semantic over-
lap in the corpus was the one between contrast and temporality. A number of
markers in both English and French – such as *still, while, at the same time; alors
que, tandis que, en même temps* – had the potential to function either as conjunc-
tive markers of contrast – as in Examples (8) and (9) – or as temporal markers – as
in (10) and (11).

(8)   Japan's economy has fared no better than Germany's in the past six quarters
        in real terms. During the past 12 months it has barely grown at all. **Still**, in
        nominal terms, all important for the moment, the economy has had its best
        run in decades.   (Mult-Ed-EN – The Financial Times)

(9)   L'automobiliste ne pollue que par nécessité, il n'appuie pas sur l'accélérateur
        pour le plaisir d'envoyer du gaz carbonique dans l'air. Il est presque
        innocent, **tandis que** le fumeur, dès qu'il allume une cigarette, veut
        entraîner les autres dans sa propre mort.   (Mult-Ed-FR – Libération)

(10) Skilled in war but also in the arts of peace, he was a seaman, naval officer,
        farmer and administrator whose vision of what Australia could become is
        still an inspiration.   (Mult-Ed-EN – The Guardian)

(11) On a déjà dit quels jugements à la fois espiègles, impertinents, et en
        tout cas bien libres elle formulait sur les grands de ce monde (Churchill,
        Krouchtchev, de Gaulle, Malraux), **tandis qu'**elle les rencontrait à la Maison
        Blanche ou dans les capitales du monde aux côté de son mari.

                                                            (Mult-Ed-FR – Le Nouvel Observateur)

While in the examples above, the delimitation between the temporal and the con-
trastive uses of these markers appears to be rather clear, for a number of occurrences

in the corpus, it was much more difficult to clearly draw the line between the two meanings (see also among others: Miltsakaki et al. 2005; Meyer 2011; Meyer et al. 2011; Bilger & Cappeau 2013 on the ambiguity between temporal and contrastive uses of CMs). Such ambiguous instances – which may be said to exemplify Petukhova and Bunt's (2009) notion of 'simultaneous multifunctionality' defined above – are provided in (12) to (14).

(12)   Deirdre Kelly was shown doing the best she could for her kids in straitened circumstances, stowing money for local addicts who knew they couldn't trust themselves. **While** others rushed to judge, she switched on the kettle, and was even heard discussing the roots of one neighbour's addictions in childhood sexual abuse.   (Mult-Ed-EN – The Guardian)

(13)   Just as a dragon sporting a three-piece suit may **still** feel and act like – and be perceived locally as – a dragon, a Leninist state sporting a legal system may **still** behave like and be understood locally as a Leninist state.
                                                        (LOCRA-EN – Sociology)

(14)   Un exécutif spectaculairement affaibli, pris en tenailles entre des syndicats d'un autre âge et une gauche jusqu'au-boutiste. Voilà à quoi ressemble la France de ce mois de juin 2014. On croit y entendre le tocsin. **Tandis qu**'autour de nous des pays se réforment à marche forcée, la France se rejoue un film aux couleurs sépia.   (Mult-Ed-FR – Le Figaro)

In (12), *while* may be interpreted as both: (i) a signal that Deirdre Kelly's behaviour *contrasts* with other people's; and/or (ii) an indication that her reaction occurred *at the same time as* other people's. Note that the ambiguous character of *while* has already been underlined in a number of previous discourse studies: for example, it has led the annotators in Miltsakaki et al.'s (2005) sense annotation experiment to label no less than 10% of the occurrences of *while* in their corpus as 'uncertain', while disagreeing on another 13% of the cases. The same type of ambiguity is observed in Example (14), where *tandis que* may be said to express either (i) a contrast between France's conservative policies and the reforms taking place in the neighbouring countries; or (ii) a relation of simultaneity between S1 and S2 – hence being interchangeable with the temporal phrase *pendant que*. Likewise, in (13), the two occurrences of *still* may be interpreted in two different – though perhaps complementary – ways: (i) *still* can be viewed as a temporal marker, which indicates that dragons would *continue* to be perceived as such if they wore suits, and that Leninist states would *continue* to be perceived as such if they sported a legal system; or (ii) *still* can be read as a contrastive marker, indicating that *despite the fact that* a dragon is wearing a suit, it is perceived as a dragon all the same; or that *despite the fact that* they sport a legal system, Leninist states are nonetheless perceived as Leninist states. With respect to *still*, it must also be noted that the

ambiguity between contrast and temporality mainly arose with medial uses of the marker. When *still* was used initially, as in (8) above, it most often conveyed a very clearly contrastive meaning. Some of its medial uses, by contrast, were much more ambiguous: alongside very clear cases of temporal uses such as (10), a number of medial occurrences appeared to express both contrast and temporality simultaneously, as in (13). Thus, in this case, there seems to be a correlation between the position of the marker and its function in discourse (see also Greenbaum 1969; Martin 1992: 43; Degand & Fagard 2011 for similar observations).

Another set of particularly problematic instances in the corpus pertained to markers that appeared to combine a meaning of contrast with one of addition. Such cases, which were especially frequent with *while*, but also common with French *tandis que, or, mais,* or with English *but*, are exemplified in (15) to (18) below.

(15)   Whatever the cause of these "extreme events", the UK is not alone in experiencing nature's wrath. Greenland and Norway are exceedingly warm; California is experiencing a drought; Australia has weathered its hottest year on record, **while** many states in the US are experiencing snowfall in "double digits" […].   (Mult-Ed-EN – The Observer)

(16)   This is the consequence of Vladimir Putin's decision last Saturday in effect to annex the Crimean peninsula in a show of anger at the toppling of his ally Viktor Yanukovich. In response, Barack Obama and EU leaders have launched a first salvo of penalties designed both to express their unhappiness and to impose costs on Moscow. So far, the sanctions are modest. Washington is imposing travel bans **while** Europe's response has been largely symbolic.   (Mult-Ed-EN – The Financial Times)

(17)   A croire les déclarations à droite mais aussi à gauche, à lire les manchettes de nos confrères, la France est confrontée à une menace rom. Un journal qui se dit libéral a dénoncé hier, carte à l'appui, «la France des campements roms», **tandis qu'**un hebdomadaire d'extrême droite sombrant dans le caniveau a titré «Roms l'overdose» sondages à l'appui.
                         (Mult-Ed-FR – Libération)

(18)   Il y a en effet plusieurs sortes de tutoiements. Celui cordial ou amical, que j'emploie volontiers, comme tant de journalistes. J'ai été militant du PS, parti dans lequel le tutoiement est de règle, ce qui est plutôt agréable. **Mais** il y a aussi le tutoiement agressif, tel celui qu'emploient les automobilistes quand ils s'invectivent.   (Mult-Ed-FR – Le Nouvel Observateur)

In Examples (15) and (16), the relationship conveyed by *while* lies at the border between addition and contrast. On the one hand, the two segments related by *while* are both oriented in the same argumentative direction, viz. in (15), S1 and

S2 both contribute to corroborating the claim that *the UK is not alone in experiencing nature's wrath*; whereas in (16), both segments illustrate the assertion that *the sanctions of the West towards Russia have so far been modest*. On the other hand, in both examples, *while* also underlines differences between the related segments: in (15), S1 provides examples of regions that have suffered from particularly hot weather, whereas S2 focuses on areas which have experienced unusually cold weather as a result of 'nature's wrath'. Similarly, in (16), the CM compares the type of sanctions imposed upon Putin by the leaders on each side of the Atlantic. A very similar situation is observed in Example (17): on the one hand, French *tandis que* links two arguments that point in the same argumentative direction, with each one exemplifying the statement that the French media have been denouncing the threat posed by the Roma population on the country. On the other hand, *tandis que* also compares the ways in which this has been done by a liberal and by a far-right newspaper. Finally, in Example (18), the coordinator *mais* could also be said to perform two functions in discourse. On the one hand, it signals a contrast between the two types of *tutoiements* described by the writer, viz. the friendly one and the aggressive one. On the other hand, when considered together with the adverb *aussi* that follows it, this use of *mais* also shares some traits with the additive *mais* described by Anscombre and Ducrot (1977) or Van De Voorde (1992): in this specific case, *mais* also signals the second stage in an enumeration initiated by the first sentence in the example (*Il y a en effet plusieurs sortes de tutoiement*). Note that the proximity between the discourse relations of addition and contrast may seem surprising at first sight: for example, these two meanings are presented as opposites of each other in Sanders et al.'s (1992) Cognitive approach to Coherence Relations, where contrast is said to belong to the category of negative relations (as it involves the negation of some of the content presented in the related segments), whereas addition is viewed as a positive relation (see Chapter 2 for more on this). On the other hand, it is interesting to underline that the fuzziness of the boundary between addition and contrast is partly reflected in the fairly high frequency of the additive sequences *not only… but (also), non seulement… mais (aussi)* and their variants, where the most central marker of contrast in each language (viz. *but* and *mais*) is in fact used to express a meaning relation of addition, as in (19) and (20). Such additive uses of *mais* and *but* were of course weeded out from the data set.

(19)   Moreover, by making it a legal requirement to spend a rising amount of money on aid, Whitehall officials will be breaking the law if they fail to find causes to give it to. **Not only** is this inefficient, **but** it increases the chances that money will end up in the wrong hands, as it almost certainly has already.   (Mult-Ed-EN – The Telegraph)

(20)   Autrement dit, **non seulement** Meyssan trompait les millions de
       téléspectateurs qui l'ont vu déblatérer sans contradicteur à l'invitation de
       Thierry Ardisson – jour noir dans l'histoire de la télévision française –
       mais il insultait **aussi** les familles des victimes, frappées par un deuil aussi
       absurde que cruel pour être ensuite niées et salies par un escroc intellectuel
       sans foi ni loi.   (Mult-Ed-FR – Le Nouvel Observateur)

Finally, alongside the numerous occurrences of CMs combining meanings of
either (i) contrast and temporality; or (ii) contrast and addition, the process of
semantic disambiguation also raised more marginal issues, which were more spe-
cific to one or a few markers in particular. In French, for example, the CAs *plutôt,
tout de même* and *quand même* may all be used either as CMs of contrast, or as
emphatic markers. In (21), for instance, *plutôt* expresses a meaning of correction,
viz. it signals that the information presented in S2 provides the correct alternative
to the information that was negated in S1. In (22), by contrast, *plutôt* does not
perform a conjunctive function: rather, it functions as an intensifying adverb that
modifies the adjective *inquiète* (*worried*), and could be translated as *rather, fairly*
or *quite*.

(21)   Ce n'est pas que l'Allemagne ou la France manquent d'énergie ou d'ambition
       pour l'Europe – bien au contraire. C'est **plutôt** que leur conception de
       l'Europe devient minoritaire au sein de l'Union.   (Mult-Ed-FR – Le Monde)

(22)   Enfin, le pessimisme au regard de l'avenir professionnel est frappant car près
       de la moitié des enseignants (43 %) est **plutôt** inquiète ou très inquiète à cet
       égard, particulièrement chez les enseignants en fin de carrière embauchés
       avant les années 1980 (49 %) et ceux ayant amorcé leur carrière en 1990–
       1999 (46 %) ainsi que chez les enseignants du secondaire (47 %).
                                                       (LOCRA-FR – Education)

Although in the examples above, the distinction between the contrastive and the
emphatic uses of the marker appears fairly clear-cut, a number of instances in
the corpus data were found to lie at the boundary between these two functions.
In (23), for instance, *plutôt* may be interpreted in two different ways. On the one
hand, it may be said to express a relation of contrast (and, more precisely, oppo-
sition) between the political orientation of the liberals and that of the interven-
tionists. As such, *plutôt* reinforces the meaning of the subordinator *alors que* that
precedes it. Alternatively, *plutôt* may also be viewed as an intensifier, which quali-
fies the phrase that follows: it suggests that interventionists *tend to* stand towards
the left of the political spectrum.

(23)   [L]es libéraux se sentent beaucoup plus souvent de droite alors que les
       interventionnistes sont **plutôt** à gauche.   (LOCRA-FR – Political science)

In summary, the examination of the corpus data shed light on the ambiguity of the meaning relations conveyed by some CM tokens, as well as the permeability of the boundaries between the various semantic categories of discourse relations identified in the literature. As a result, it was sometimes very difficult to determine whether or not a given occurrence qualified as a contrastive CM and should be included in the data set. The strategies adopted to deal with such indecisions in the most reliable way possible are described in Section 6.1.3. Before turning to that aspect, however, it also interesting to underline that another element that came out from the semantic disambiguation of the data was the utmost importance of shared knowledge for the interpretation of conjunctive markers. A large number of the texts in the corpus – e.g. most notably, the research articles making up the LOCRA subcorpus – conveyed highly specialised ideational content. This rendered the disambiguation task significantly more complex: as I sometimes struggled to fully understand the ideational content expressed in the two segments related by the CM, it was very difficult to determine what the most justified interpretation of a given CM in a specific context could be. In Example (24) below, for instance, one has to know that fracking (i.e. a drilling method aiming to release gas out of shale rock as a source of energy) is a new, exploratory method which has not yet been implemented in Britain, to interpret *still* as a fully contrastive instead of a temporal marker. Likewise, to correctly interpret the marker in (25), the analyst has to know that Matteo Renzi is a newly elected politician, heading a newly-formed coalition, so that it makes little sense to interpret *still* as a temporal marker.

(24) While the public will have concerns, which need to be aired, appreciated and addressed, fracking **still** has the capacity to work wonders for Britain.
(Mult-Ed-EN – The Telegraph)

(25) The question now is whether Mr Renzi can translate this victory into a boost for his reform agenda. He has a highly ambitious manifesto, seeking to overhaul the rigid Italian labour market, cut public sending, redesign the parliamentary system and change the electoral law. But in spite of the election victory, he **still** heads a complex coalition that relies on centre-right parties and a fractious parliament to get things done.
(Mult-Ed-EN – The Financial Times)

The importance of contextual information for discourse interpretation has also been underlined by Spooren and Degand (2010: 255), who noted that different discourse analysts may interpret the context of a CM in different ways, which may in turn lead to disagreements and subjectivity in the coding of discourse relations. While I did my best to minimise the cases of misinterpretations due to my own lack of background knowledge (e.g. by seeking relevant information on the topics covered whenever I felt that I missed some of the elements necessary to properly

interpret a given CM token), it is of course still possible that another analyst with more extensive knowledge in political science, sociology or psychology, for instance, might have provided a partly different interpretation of some of the CMs in the corpus. Even though a certain amount of subjectivity has been acknowledged to be inevitable in discourse analysis (see e.g. Craggs & McGee Wood 2005: 293; Spooren & Degand 2010: 263; Crible 2017: 263 on this), it is nonetheless an aspect that needs to be borne in mind when interpreting the corpus results.

6.1.3   Dealing with ambiguity: The use of double tags

In the previous section, it was demonstrated that CMs do not always unambiguously signal one single, clearly-identifiable discourse relation, but that a number of markers convey meanings that lie at the boundary between two (or more) semantic categories, such as contrast and addition, or contrast and temporality. Such ambiguities can constitute a true challenge for the analyst, who may find it difficult to decide on the most appropriate semantic label to be assigned to a given CM token. A number of strategies may be used by the researcher to try to resolve such ambiguities: for example, they may attempt to substitute the ambiguous marker by other, more unequivocal ones (see e.g. Zufferey & Degand 2017: 13): in (26), for instance, *still* is successfully substituted by *even so*, which confirms that in this context, the marker conveys a concessive meaning. The researcher may also try to paraphrase the meaning conveyed by the marker – as in (27) – or translate it into another language – as in (28) (see e.g. Sanders 1997; Cartoni et al. 2013 on the value of paraphrases and translations as tools for disambiguation).

(26)   Japan's economy has fared no better than Germany's in the past six quarters in real terms. During the past 12 months it has barely grown at all. **Still [Even so]**, in nominal terms, all important for the moment, the economy has had its best run in decades.
>   (Mult-Ed-EN – The Financial Times; example of a substitution test)

(27)   **Despite the fact that** Japan's economy has been sluggish in the past year, it has had its best run in decades.
>   (paraphrase test performed on Example (26))

(28)   Japan's economy has fared no better than Germany's in the past six quarters in real terms. During the past 12 months it has barely grown at all. **Still [Malgré tout]**, in nominal terms, all important for the moment, the economy has had its best run in decades.
>   (translation test performed on Example (26))

For truly ambiguous CM uses, however, these tests are not necessarily helpful. In such cases, several substitutions, paraphrases and/or translations are usually equally plausible and acceptable, as can be seen in Example (29), where *still* may be

translated into French by either the concessive CM *quand même* or the temporal adverb *toujours*. This was also visible in many examples from the previous section, which could frequently be paraphrased in two perfectly plausible ways. In other cases, the best translation of a given CM is equally ambiguous as the original, as in Examples (30) (see also Cartoni et al. 2013: 70 on this).

(29)   It is destined to be a long winter, and a predictable one. On the basis of the opening "Test", this Ashes series will be no more of a contest than any of the previous battles had proved for the past 16 years. This is hardly entertainment; even the Australians will (eventually) tire of winning this easily and gloating about the plight of the pathetic Poms. Sport requires a sense of drama, a real struggle, some doubt as to the final outcome. None of this exists at the moment. The situation is not, however, irretrievable. The adoption of a simple, if admittedly radical, ten-point plan could serve to make the next four Tests a little more competitive. (1) If England win the toss then they should be permitted a second chance to reconsider whether they might bat or field after the conclusion of the first day's play. Another opportunity to revisit the issue might be helpful during drinks on the fourth afternoon (if the match lasts that long). (2) In the spirit of Don Bradman, who apparently honed his technique in a similar fashion, Matthew Hayden should now be obliged to come to the crease holding only a stump while the wicket should consist of three bats neatly lined up together. (3) If Hayden **still [quand même|toujours]** scores more than the entire England team combined then he must be forced to bat using one of the bails while the sightscreen should serve as the wicket.   (Mult-Ed-EN – The Times)

(30)   Precisely because the NHS is free at the point of use, we have no enforceable rights as patients and have to accept what we are given. The result is an uncivilised system in which those who can pull strings or threaten legal action get their way, **while [tandis que]** the majority simply suffer in silence.
                    (Mult-Ed-EN – The Telegraph)

What examples such as these show is that in some cases, the meaning of a given CM is simply inherently ambiguous: two or more meanings coexist in a single CM occurrence, so that it is not really possible – or indeed desirable – to place them in a well-defined semantic category such as 'contrast', 'addition' or 'causality'. Yet, as pointed out by Cartoni et al. (2013: 79), discourse annotators are usually tempted to choose one single label or category among a range of possibilities. The tendency to 'idealise' the data in quantitative corpus linguistics was also highlighted by McEnery and Wilson (2001: 77):

> Quantification […] entails classification. For statistical purposes, these classifications have to be of the hard-and-fast (so-called 'Aristotelian') type, that is an item either belongs in class *x* or it doesn't […]. In practice, however, many linguistic items and phenomena do not fit this Aristotelian model: rather, they are

> consistent with the more recent notions of 'fuzzy sets', where some phenomena may clearly belong in class *x* but others have a more dubious status and may belong in potentially more than one class […]. Quantitative analysis may therefore entail in some circumstances a certain idealisation of the data: it forces the analyst to make a decision which is perhaps not a 100 per cent accurate reflection of the reality contained in the data.

In order to avoid such a deceptive idealisation of my corpus data, and following researchers such as Miltsakaki et al. (2005), Meyer et al. (2011), Cartoni et al. (2013) or Zufferey and Degand (2017), I have chosen to allow for double semantic tags, so as not to be forced to choose between categories when two meanings genuinely coexisted in a single CM occurrence. Thus, the markers which appeared to express meanings of both contrast and temporality were labelled 'temporal-contrastive', while those which simultaneously conveyed meanings of contrast and addition were labelled 'additive-contrastive'. Members of both these 'hybrid' categories were kept in the data set. This methodological choice was felt to be the one which best reflected the linguistic reality emerging from the corpus data, where a range of CM uses were found to be truly polyfunctional. Accordingly, attempting to force them in either one or the other category would have provided an oversimplified picture of CM use, masking the complexity of the relations that these items may express. In addition, as pointed out by Cartoni et al. (2013: 79), systematically trying to choose between labels would have left the door open for arbitrary decisions.

   Importantly, however, it must be noted that the use of these borderline categories did not magically solve all the problems posed by semantic disambiguation. As is often the case in semantics, the distinctions between temporal and contrastive meanings, and between the additive and the contrastive categories, are best represented on a continuum from 'mostly contrastive' to 'mostly temporal/additive'. This means that, among the CMs which expressed both relationships simultaneously, some appeared to occupy some sort of middle ground between two semantic categories, but others were especially contrastive with a taint of temporality/addition, and yet others were primarily temporal or additive, while displaying a very light contrastive flavour. In Example (31), for instance, even though there is certainly a partial contradiction between the action of *standing by* and the fact that more than 1,000 people were being killed, the dominant function of *while* in this particular case is to signal the concurrence of the two events being described. As a result, it was not always easy to draw the line between the CMs which could be considered as true markers of contrast, and those whose contrastive meaning was too secondary to justify their inclusion in the data set. Another set of – rather recurrent – cases which were particularly difficult to deal with pertains to markers which may express either contrast or temporality, and co-occurred with other

temporal and contrastive markers, as in (32). In these cases, it was very hard to determine whether the CM was meant to reinforce the contrastive relation (as expressed by *yet*) or the temporal meaning (conveyed by the verb *remains*).

(31)   Only a few years ago, he appeared unelectable as prime minister. A cloud hung over him from revenge killings in Gujarat in 2002 when, as the state's chief minister, he was accused of standing by **while** more than 1,000 people, mostly Muslims, were killed.   (Mult-Ed-EN – The Financial Times)

(32)   The concept of inequality has been extensively studied <u>yet</u> it **still** <u>remains</u> contentious.   (LOCRA-EN – Sociology)

One strategy which I used to remain as consistent as possible in the decisions made was that, for the very ambiguous CMs (e.g. *while, still, alors que, tandis que*), I first worked on small samples of randomly selected tokens. The objective was to come up with a dependable 'disambiguation strategy' to be applied to the entire data set (for example, by establishing that each time *still* co-occurred with both a temporal and a contrastive marker, it would be labelled 'temporal-contrastive'). This decision was inspired by Spooren and Degand's (2010: 251) observation that the quality of the semantic coding of discourse relations tends to improve with practice, so that "[c]omplex coding […] requires a warming-up phase".

In addition to isolating the markers expressing a meaning of contrast, the identification of the set of conjunctive markers in my corpus also required that I select only the items which performed a genuine discourse function, i.e. the markers linking segments corresponding to clauses or larger units. While such a procedure – usually referred to as 'discourse segmentation' – may seem rather straightforward at first sight, it raises a series of very challenging issues, especially when approached from a contrastive perspective. These are discussed in the following section.

6.2   Syntactic segmentation

6.2.1   What do conjunctive markers link?

Extensive research has been devoted to identifying the number and types of relations that may hold between two segments of discourse. Yet, although the identification of the units between which such relations hold is a prerequisite for the identification of the discourse relations themselves, far less attention has been paid to determining the exact nature and delimitation of the segments that may be linked by means of conjunctive markers (but see Schilperoord & Verhagen 1998; Verhagen 2001; Hoek et al. 2018 for notable exceptions). As pointed out in Chapter 2, in research on discourse relations, it is generally accepted that the

minimal units which may be related by means of a CM correspond to clauses. This means that – at least in written language – a given marker is considered to perform a discourse function when it links two clauses (or larger units) – as in (33) – whereas a linking expression relating segments smaller than clauses – as in (34) – is not viewed as a conjunctive marker (see e.g. Mann & Thompson 1988; Carlson & Marcu 2001; Miltsakaki et al. 2004).

(33)   Manufacturers should be honest about sugar content, **and** consumers should better inform themselves about how much is in what and the ill effects of consuming too much of it.   (Mult-Ed-EN – The Telegraph)

(34)   Germany **and** Italy could be hurt by restrictions on Russian oil and gas imports.   (Mult-Ed-EN – The Financial Times)

In addition, a number of researchers have in fact included some more specific reflections on segmentation issues, but these have mostly consisted in identifying a range of exceptions to the basic 'clause principle'. Mann and Thompson (1988: 248), for example, consider all clauses "except […] clausal subjects and complements and restrictive relative clauses" to function as segments of discourse. Likewise, although Miltsakaki et al.'s (2004) segmentation procedure is essentially reliant on the notion of clause, they also include (i) nominal phrases expressing an event or state; and (ii) discourse deictics expressing an event or state within the set of discourse units. Carlson and Marcu (2011), for their part, exclude subject and complement clauses, but include phrasal expressions introduced by a 'strong' discourse cue (such as *despite, as a result of, because*, etc.).

However, as pointed out by Muller et al. (2012: 4), little attention has been devoted to the notion of 'clause' itself, which is usually taken for granted in current discourse research. More specifically, the criterion used to grant clausal (and thus, discourse) status to a given textual segment is commonly limited to whether or not it contains a verb. Thus, for instance, from the mere fact that it does not contain a verbal element, Carlson and Marcu (2001: 28) consider the segment introduced by *while* in Example (35) to be a non-clausal unit. Accordingly, *while* is not viewed as a conjunctive marker in this example.

(35)   But the technology, **<u>while</u>** reliable, is far slower than the widely used hard drives.   (Carlson & Marcu 2001: 28)

This overreliance on the verb tends to presuppose that the clause is a straightforward concept, which is far from being the case. Among other problems, the verb-based conception of the clause totally obscures the fact that clauses are not always complete but may involve the ellipsis of one or more of their constituents – including the verb phrase (see also Hoek et al. 2018: 362 on this). Therefore, the presence of a verb is not a sufficient criterion to discriminate between clausal and

non-clausal (and thus, discourse and non-discourse) segments. This is especially true when working in a contrastive perspective since, as we shall see, the necessary ingredients to form a clause may partly differ from one language to another (see also Hansen-Schirra & Neumann 2003: 2 on the challenges posed by textual segmentation across languages). In the remainder of this section, I explore issues of discourse segmentation for the cross-linguistic analysis of conjunctive markers in written language. Following common practice, the clause is taken to be the minimal unit of discourse. However, the main contribution of the present chapter is that, instead of relying on a fairly vague definition of clause, then supplementing it with a range of exceptions and addenda, it attempts to redefine the boundaries of the clause in a rigorous and transparent fashion, while also making sure of the validity of this definition for the description of both English and French.

6.2.2 Clauses in English and French

The problems arising from a fully verb-based definition of the clause appear clearly when comparing the descriptions of this concept provided in the English and the French literature. It is true that in English, the definition of the clause is largely reliant on the presence of a verb. According to Thompson (2014: 17), for example, a clause can be defined as "any stretch of language centred around a verbal group". Likewise, for Carter and McCarthy (2006: 486), clauses are made up of two subparts, viz. a subject and a predicate, which is itself composed of "a verb phrase and any other accompanying elements". For them, too, the clause can be said to "centre around a verb phrase" (ibid.). A similar view is also taken by Biber et al. (1999: 120), who define the clause as "a unit structured around a verb phrase". Accordingly, Biber et al. (ibid.: 224) refer to segments such as the one in (36) as "non-clausal material".

(36)   And now for something completely different: cheap and cheerful claret.
(Biber et al. 1999: 224)

One type of verbless segment that nonetheless appears to be granted clausal status in English grammars pertains to hypotactic clauses with an ellipted copular verb, as in (37) and (38) (see e.g. Quirk et al. 1985: 992; Johansson & Lysvåg 1986: 175; Biber et al. 1999: 201). In addition, Huddleston and Pullum (2002: 944–945) have used the term 'minor clause' to refer to a very specific set of marginal syntactic structures, some of which verbless, which may nevertheless be viewed as clauses. More specifically, Huddleston and Pullum (ibid.) grant the status of 'clause' to both verbless directives – as in (39) – and verbless parallel structures – as in (40). In all the other contexts, by contrast, the presence of a verb truly seems to be a defining feature of the English clause.

(37)   <u>Although always helpful</u>, he was not much liked.   (Quirk et al. 1985: 992)

(38)   Wagner, <u>though a tremendous genius</u>, gorged music, like a German who overeats.   (Johansson & Lysvåg 1986: 175)

(39)   All aboard!   (Huddleston & Pullum 2002: 945)

(40)   Out of sight, out of mind.   (ibid.)

In the French literature, on the other hand, the definition of the clause appears to be rather less dependent upon the notion of verb than in English. In French grammars, the definition of the clause (or sentence)[1] usually revolves around the presence of a predicate rather than a verb, where 'predicate' refers to 'what is being said about the subject of the clause' (see e.g. Le Goffic 1994: 8; Grevisse & Goosse 1995: 76; 2011: 222).[2] Although French grammarians generally insist that the predicate most often contains a verbal element, there is wide recognition among them that this predicate may also be verbless, whether the clause is hypotactic – as in (41) – or stands on its own – as in (42). Accordingly, verbless structures are granted far more attention than in the English literature: it is common for grammars of French to dedicate a complete section or chapter to verbless clauses and sentences (cf. e.g. Le Goffic 1994: 509–523; Grevisse & Goosse 2011: 514–516), while Lefeuvre (1999) even devoted an entire book to the description of French verbless sentences.

(41)   <u>Bien que philosophe</u>, M. Homais respectait les morts.

                          (Grevisse & Goosse 2011: 1476)

(42)   Contre la fatigue, un seul remède : le repos !   (Le Goffic 1994: 520)

Given the apparent importance of verbless clauses in French, a contrastive study that would only look at the markers linking two verbal units would likely overlook some relevant uses of French conjunctive markers. In an example such as (43), for instance, the French marker *donc* would be considered to perform the function of conjunctive marker, since the segment that it introduces is considered as a clause in its own right (see Danlos et al. 2015b: 6) – instead of being dismissed as 'non-clausal material', as would be the case in English. Yet, a fully verb-based definition of the clause would lead me to discard such examples as irrelevant.

---

1.   It is worth noting that, whereas English grammars tend to focus their description on the clause, in French the basic unit of syntactic description is commonly the sentence.

2.   Importantly, as made clear by Lefeuvre (1999: 31), this subject need not be explicit. For example, if someone were to utter the word 'Délicieux!' while tasting a pastry, it would be very clear that the implicit subject over which something is being stated is the pastry being eaten by the speaker.

(43)   Luc a dû démissionner. Une véritable tragédie **donc** pour lui.

(Danlos et al. 2015b: 6)

Therefore, in order not to bias the analysis in favour of English – and in line with the concerns of comparability which underlie most of the theoretical decisions made in the present study – it was necessary to adopt a fairly broad definition of the clause, which allowed for the inclusion of verbless and elliptical segments, hence doing justice to the specificities of the French linguistic system. Accordingly, instead of being defined as a structure centred around a verb, in this book, the clause is defined as 'a structure which is minimally constituted of a predicate' – i.e. an element whose core function is simply to provide some information about another entity in discourse (see Lefeuvre 1999; Hoek et al. 2018: 363 for similar definitions). In most cases, the predicate around which the clause is built is verbal: it may appear either in finite (or tensed) form – as in (44) and (45) – or in non-finite (i.e. infinitive or participial) form – as in (46) and (47). In other cases, however, the predicate may also be verbless, as in Examples (48) to (50).

(44)   Compared to their female counterparts, Finnish male teachers had significantly higher evaluations of their competence in dealing with behaviour issues, while such a connection was not found in the Chinese and South African samples. [**Nonetheless**, studies from other countries have reported a connection between gender and efficacy in classroom management].   (LOCRA-EN – Education)

(45)   Puisque nos données sont constituées d'un nombre relativement important de points temporels, nous testons systématiquement la présence d'une tendance linéaire. [**Cependant**, nous représentons toujours dans les figures les coefficients annuels. (non linéaires)], puisqu'ils permettent mieux d'évaluer les tendances ainsi que l'homogénéité des séries d'enquêtes

(LOCRA-FR – Sociology)

(46)   On occasion, [**though** without ever condoning the killing], it is even possible to sympathise with what drove them to it.

(Mult-Ed-EN – The Guardian)

(47)   Devant la levée de boucliers des présidents d'université, des directeurs de grande école et des chefs d'entreprise, mais aussi devant l'incompréhension des pays étrangers concernés, M. Guéant avait fini par amender son texte, [sans **toutefois** y renoncer].   (Mult-Ed-FR – Le Monde)

(48)   [**Although** without a throne of their own], the French love the style that monarchy brings with it.   (Mult-Ed-EN – The Telegraph)

(49)   Lui aussi, [**quoique** homme libre et indifférent à l'opinion des autres], ne voulait pas handicaper sa présidence en choquant une partie de l'opinion.

(Mult-Ed-FR – Libération)

(50)   La nervosité de Valérie Lecasble, directrice générale d'i-Télé, après cette mise en ligne, montre bien qu'elle s'y est sentie, quelque part, dévoilée. [**Pourtant**, en première lecture, rien d'inavouable].

(Mult-Ed-FR – Libération)

Another important distinction with respect to the description of clauses pertains to the rank status that they may occupy in the linguistic system. More specifically, clauses may be either independent, or syntactically dependent on another clause. In Examples (51) and (52), the clauses hosting the CMs function as independent – or 'main' – clauses, which means that they are not structurally dependent on any other element, but may stand on their own in discourse (Carter & McCarthy 2006: 487). Note that according to Carter and McCarthy (ibid.: 544), main clauses *must* include a verb in the finite form in order to be grammatically correct. Yet, as we saw above, the literature on French commonly discusses the possibility of verb-less clauses standing on their own in discourse. The rank status of such verbless structures is further discussed below.

(51)   Moreover, the predictability of response cues decreased the estimated 1/f exponents in response times toward random, white noise fluctuations.
By contrast, 1/f exponents in key-contact durations were unaffected
by cue predictability.   (LOCRA-EN – Psychology)

(52)   Contrairement à ce qui s'est passé en Angleterre, ces données mettent en évidence le maintien d'une forte synchronisation des repas. **Cependant** les structures de repas se sont transformées en France.

(LOCRA-FR – Anthropology)

Other clauses, by contrast, *cannot* stand on their own in discourse, but are syn-tactically dependent on another clause. Unlike main clauses – which necessarily contain a verb in the finite form – dependent clauses may be either finite, non-finite or verbless. In Systemic Functional Linguistics, two main types of dependent clauses are usually identified, viz. hypotactic and embedded clauses. In order to clearly set out the difference between those two clause types, it is necessary to resort to the SFL notion of 'rank scale', which is based on the idea that any mean-ingful linguistic unit at one rank of the language system can be split into units of the rank below on the following scale: clause > group > word > morpheme (see e.g. Bloor & Bloor 2004: 7–8; Thompson 2014: 21). Even though they may not stand on their own, hypotactic clauses function at the 'clause' rank of the linguistic system: they combine with other, independent clauses to form clause complexes (see Chapter 2 for a definition of clause complex). Hypotactic clauses may stand in a range of semantic relationships with the main clause to which they are attached: they can provide various types of circumstantial information (e.g. temporal, causal, conditional, concessive, etc.) on their main clause – as in Examples (53)

to (55); or add extra information about one subpart of the clause to which they are attached, thus performing the function of non-defining relative clauses, as in (56) and (57). In addition, Systemic Functional linguists also include dependent clauses that are used to quote or report speech or ideas (viz. so-called 'projection clauses', exemplified in (58) and (59)) within the category of hypotactic clauses (see e.g. Butt et al. 2000: 167–168; Halliday & Matthiessen 2004: 443–444; Thompson 2014: 201–206).

(53)   If the Church is as keen on diversity as it says it is, the two Primates should not sing from the same political hymn-sheet. <u>If, **on the other hand,** Dr Sentamu shifts his gaze elsewhere</u>, he could achieve a great deal.
(Mult-Ed-EN – The Telegraph)

(54)   Si la politique était une science exacte, Jacques Chirac ferait tranquillement de Nicolas Sarkozy son héritier. <u>Comme elle est **au contraire** un art martial</u>, les deux plus fortes personnalités de la droite vont tout faire pour s'entre-tuer.
(Mult-Ed-FR – Libération)

(55)   <u>**Même s**'il ne faut pas crier au loup quand se présente un simple furet</u>, il est nécessaire de mettre rapidement fin à cette dérive.
(Mult-Ed-FR – Le Figaro)

(56)   Gilligan's report was not checked and his notes (<u>which were **anyway** lamentable</u>) were not examined until a month later.
(Mult-Ed-EN – The Telegraph)

(57)   Et, pour la première fois, les Etats-Unis demandent très ouvertement la fin de l'«occupation» syrienne du Liban, <u>dont ils s'accommodent **pourtant** depuis longtemps</u>   (Mult-Ed-FR – Le Figaro)

(58)   The police say that Brown attacked the officer and was shot in a scuffle. A friend who was with Brown says <u>that, **on the contrary**, the teenager had his hands in the air and was shouting "Don't shoot" when he was killed.</u>
(Mult-Ed-EN – The Financial Times)

(59)   On déclarerait, par exemple, que les dépenses de fonctionnement doivent être couvertes par l'impôt mais <u>qu'**en revanche** le financement des investissements publics par l'emprunt - ou même par la création monétaire – est légitime</u>.   (Mult-Ed-FR – Le Nouvel Observateur)

As opposed to hypotactic clauses, embedded clauses (also referred to as 'rank-shifted' or 'downranked' clauses) do not function at clause rank in the linguistic system. Instead of combining with other, independent clauses to form clause complexes, they are 'demoted', as it were, to function as a group – or even as *part of* a group – in another clause (Halliday & Matthiessen 2004: 426–437; Caffarel 2006: 29; Thompson 2014: 24–26). Thus, in Example (60), for instance, the *that*-clause which

appears between square brackets performs the function of subject in another, independent clause. In (61), on the other hand, the defining relative clause only constitutes a subpart of the subject of the main clause: it is used to post-modify the noun phrase used as head of the group. Therefore, for Systemic Functional linguists, defining and non-defining relative clauses do not operate at the same level of the linguistic system: whereas non-defining relative clauses are viewed as hypotactic clauses functioning at clause rank, defining relative clauses are categorised as embedded clauses functioning at group rank.

(60)   [That there had soon been a reconciliation] was due to Albert.

(Thompson 2014: 25)

(61)   [The idea that this new method would bring profit] soon drew other
       manufacturers into the field.   (ibid.)

Because they do not function at clause rank, embedded clauses have been denied the status of discourse segments by a number of researchers (see e.g. Mann & Thompson 1988; Schilperoord & Verhagen 1998; Carlson & Marcu 2001). Accordingly, in these studies the markers in Examples (62) to (64) would not be granted conjunctive status.

(62)   Sci-fi fans inhabit a world that […] [they have built for themselves] **yet**
       [share with the rest of us].   (Mult-Ed-EN – The Guardian)

(63)   A cliché of business, these days, is [that a good reputation takes years to
       build up] **but** [can be lost in an instant].   (Mult-Ed-EN – The Economist)

(64)   C'était un homme [qui ressemblait à tous les autres], **mais** [qui n'était pas
       exactement comme eux].   (Mult-Ed-FR – Le Figaro)

In line with Hoek et al. (2018), however, the present study views embedded clauses as discourse segments in their own right. It considers that the type of logico-semantic relationship that holds between the clauses in the examples above is largely comparable to those that may hold between independent clauses – as attested by the fact that in many cases, it is relatively easy to simply transpose such sequences into a combination of independent clauses, as in (65) or (66). Consequently, markers relating two embedded clauses are regarded as conjunctive.

(65)   A good reputation takes years to build, **but** can be lost in an instant
                                           [adapted version of Example (63)].

(66)   Cet homme ressemblait à tous les autres, **mais** il n'était pas exactement
       comme eux   [adapted version of Example (64)].

As briefly underlined above, independent and dependent clauses are each associated with specific conditions governing their correct use. On the one hand, according to Carter and McCarthy (2006: 487), main clauses need to contain a verb in the

finite form in order to be grammatically acceptable. On the other hand, dependent clauses may not stand alone in discourse, but must combine with another clause to form a structurally complete linguistic unit. Yet, the exploration of authentic corpus data revealed that, in practice, these prescriptive rules are not systematically abided by. Firstly, as discussed above, it is not uncommon for French to resort to verbless segments standing on their own, as in (67) and (68). As a matter of fact, such structures are also found in the English data – as exemplified in (69) and (70) – although they appear to be less frequent than in French. In addition, the corpus data also displayed some instances of clauses used independently, although their main verb was in the non-finite form – as in (71).

(67)   Hier, ce fut donc la rentrée du professeur Hollande sur TF1. Avec un président dans le rôle du maître d'école, appliqué et résolu, tranchant même quand il le fallait sur la Syrie. [Moins en verve **toutefois** sur la fiscalité ou sur les promesses d'un retour à la croissance].   (Mult-Ed-FR – Libération)

(68)   Seules les insultes ont volé avant que les coups ne tombent, comme dans n'importe quelle bagarre de rue. Banalisation? [Réalisme **plutôt**].
(Mult-Ed-FR – Le Figaro)

(69)   That will further discourage Labour from saying anything pro-European before the general election and could embolden British Eurosceptics, and the Eurosceptic press, to press for a more radical agenda of disengagement. Not necessarily a body blow to Labour, [but a severe constraint **nevertheless**].   (Mult-Ed-EN – The Guardian)

(70)   After the first day of the first Test, the experts condemned England's spiritless performance, bewailed their injuries, praised the power of the Australians and wrote off the Ashes. [**But** not so fast].
(Mult-Ed-EN – The Times)

(71)   Aura-t-il convaincu ceux que le manque d'audace de Hollande désespère ? [Ou, **au contraire**, conforté ceux qui restent sensibles aux sirènes du vote utile] ?   (Mult-Ed-FR – Libération)

Finally, both the English and the French data also exhibited some instances of hypotactic clauses which were not attached to any independent clause, but rather were used on their own, as in (72) and (73). Following Cosme (2004: 213), all these non-canonical syntactic structures (viz. independent clauses with a non-finite or verbless predicate, and hypotactic clauses standing alone) are included in the category of minor clauses.[3]

---

3.   Note that Cosme's (2004) category of 'minor clauses' was itself inspired from Huddleston and Pullum's (2002: 944–945), which grouped a variety of marginal clausal construction, viz.: (i) optatives (e.g. *Long live the Emperor!*); (ii) clauses with the subordinate form (e.g. *That it*

(72) The referendum, however, is also about something that it is difficult to
price – namely exercising power in your own name. [**Even if** that sort
of power comes with new levels of responsibility].

(Mult-Ed-EN – The Observer)

(73) Sans des projets susceptibles de retenir les jeunes chercheurs, elle risque fort
de se réduire à la sécheresse d'une structure bureaucratique pour laquelle la
recherche reste inscrite dans la colonne des dépenses. [**Alors qu'**elle pourrait
lui donner des ailes]. (Mult-Ed-FR – Le Monde)

To summarise the descriptions that precede, the clause may be defined rather
broadly as follows:

> A clause may be defined as a structure which is minimally constituted of a PREDI-
> CATE. The predicate around which the clause is centred is most typically VERBAL
> (viz. finite or non-finite), but it may also be VERBLESS. In terms of SYNTACTIC
> RANK, four main types of clause are identified, viz. (i) main (or independent);
> (ii) hypotactic; (iii) embedded; and (iv) minor clauses.

This means that all the markers expressing a relation of contrast between seg-
ments that correspond to the definition provided above ought to be included in
the data set. The very inclusive character of this conception of the clause is ideal
to ensure an equal treatment of the two languages investigated here (e.g. by taking
into account the greater propensity of French than English to resort to verbless
clauses). However, one major problem that it poses is that, when one can no longer
rely on the presence of a verb to discriminate clausal from non-clausal units, it can
become quite difficult to clearly identify the boundary between truly clausal seg-
ments with a verbless predicate, and phrasal segments which may not be viewed
as discourse units (see also Halliday & Matthiessen 2004: 487 on this). The blurri-
ness of the border between verbless clauses, on the one hand, and mere phrases or
groups, on the other, appears clearly from the sequence of examples in (74) to (77).

(74) The Syrian army, **though** short of combat troops, has been slowly
driving back his opponents in Damascus and Homs.

(Mult-Ed-EN – The Independent)

(75) Given the high number of journeys that need to be made, such events are
mercifully rare **though** by no means without parallel

(Mult-Ed-EN – The Telegraph).

---

should have come to this*!*); (iii) conditional fragments (e.g. *If only you'd told me earlier!*); (iv)
verbless directives (e.g. *Out of my way!*); and (v) parallel structures (e.g. *The sooner, the better*).
In view of the patterns uncovered in her corpus, Cosme (2004) broadened Huddleston and
Pullum's (2002) category so as to include all forms of verbless clauses standing on their own,
rather than just verbless directives.

(76)   Justin Welby has been commendably clear, **though** tactful, on both these
        points.   (Mult-Ed-EN – The Guardian)

(77)   Like another equally impressive **though** violent game published last week,
        The Getaway: Black Monday […], GTA is a British success story.
                                                                 (Mult-Ed-EN – The Guardian)

The recognition of the segment introduced by *though* in Example (74) as clausal is
relatively straightforward: this example is very similar to those provided by Eng-
lish grammars to illustrate verbless hypotactic clauses with an ellipted copular
verb (see e.g. Examples (37) and (38) above). Likewise, the marker in Example
(77) can fairly indisputably be identified as phrasal, as it is not possible to insert a
verbal element between the subordinator *though* and the adjective *violent* that it
introduces. The syntactic status of the segments in Examples (75) and (76), by con-
trast, is less easy to determine. On the one hand, both segments may be interpreted
as clauses involving the ellipsis of a verb phrase: in both cases, the verb phrase of
the segment preceding the CM may be repeated in the segment introduced by the
marker (e.g. in (75), the complete version of the segment introduced by *though*
is in fact the following: *though they are by no means without parallel*). Techni-
cally, both segments also meet the definition of predicate, as they provide some
information on an (ellipted) subject, viz. *such events* in (75), and *Justin Welby* in
(76). On the other hand, although a verb may be reinserted in both cases, it is also
undeniable that the type of link expressed in these examples operates at a fairly low
syntactic level. In Example (76), for instance, the subordinator is used to link two
adjective phrases. In that sense, it is perhaps questionable to treat such markers
differently from a phrasal link such as the one between the two nominal subjects
of the sentence in (78).

(78)   Germany **and** Italy could be hurt by restrictions on Russian oil and gas
        imports.   (Mult-Ed-EN – The Financial Times)

In view of this, I agree with Hoek et al. (2018: 364) that allowing objects or subject
complements such as those in (75) and (76) to be considered as individual dis-
course segments is somewhat too 'liberal'. This observation in turn points to the
necessity of restricting the definition of clause provided above since, in its current
state, it does not make it possible to systematically distinguish between the mark-
ers linking true verbless clauses and those operating at the level of the phrase. This
will be done by clearly establishing what type of segments should *not* be viewed as
clausal. In other words, due to the necessity to allow for the inclusion of verbless
units within the set of discourse segments, the definition of the clause – which now
relies on the presence of a predicate rather than a verb – has become too broad to
consistently discriminate between clausal and non-clausal segments. It appeared
that the most optimal solution to this problem was to tackle the matter from the

other end, by positing criteria to clearly identify the segments that may *not* be granted discourse status. These criteria are discussed in the following section.

6.2.3   Distinguishing between phrasal and clausal segments

In this section, I present a set of four criteria which were used to consistently differentiate between markers linking phrasal units, and markers relating verbless segments with clausal status – and thus performing a conjunctive function at discourse level. These criteria were partly inspired by the literature – and more particularly Halliday and Matthiessen's (2004: 486–489) discussion of the distinction between clause complexing, and complexing at the group or phrase rank. But in addition, they were also the result of a long process of to-ing and fro-ing between the theory and the data: a first version of these criteria was tested against the corpus data, and then revised multiple times in order to accommodate remaining problematic cases, which I was not able to handle with the previous versions of the model. The final set of criteria that a given CM token had to meet in order to be discarded as phrasal is the following:

i.   The first – and most straightforward – criterion of identification of the markers operating at phrasal level is that the segment introduced by the marker must not be built around a verbal element. This implies that all the segments which are centred around a verb phrase – whatever their rank status, viz. main, embedded, etc. – are automatically viewed as discourse units.[4]

ii. Secondly, for a given marker to be discarded as phrasal, the two related segments S1 and S2 must perform the exact same function in the clause or sentence – as in (79) and (80), where S1 and S2 both function as subject complements. By contrast, when S1 and S2 perform distinct functions in the sentence, S2

---

4.   Note that phrasal segments may nevertheless *include* a verb phrase. For example, it is common to find markers uniting two noun phrases, one (or both) of which is 'elaborated' (to use SFL's terminology) by means of one (or more) relative clause(s). This is what we observe in the following example, where the head of the second segment linked by the coordinator *but* (viz. the noun *operation*) is further specified through the use of a non-defining relative clause: *[T]his latest robbery was not [a comedy]* **but** *[a ruthless, well-executed operation in which two families were held hostage for over 24 hours]* (Mult-Ed-EN – The Guardian). However, the segment as a whole is not considered to be clausal, as it is built around the nominal head *operation*, rather than the verb phrase *hold*. This contrasts with an example such as the following, for instance, where the units that are being contrasted by means of the marker *but* are the embedded, non-defining relative clauses themselves, with each of them being centred around a verb: *There are few actions [that may be done in despair]* **but** *[which still reflect a faith in the future]* (Mult-Ed-EN – The Guardian).

is viewed as a clausal segment whose verb phrase has been ellipted (see also Halliday & Matthiessen 2004: 487 on this). Thus, for example, the subordinator in (81) is viewed as a conjunctive marker, since the second segment that it links functions as a causal adjunct, whereas the preceding segment has the function of subject complement. The coordinator in (82) is also included in the set of conjunctive markers, since the second segment in the relation functions as a manner adjunct, whereas the element preceding the CM is a subject complement.

(79)   Justin Welby has been [commendably clear], **though** [tactful], on both these points   (Mult-Ed-EN – The Guardian).

(80)   Tout à l'opposé de la vision du monde qu'ont les deux dirigeants européens les plus favorables à la levée de l'embargo sur les armes, Jacques Chirac et Gerhard Schröder. Pour eux, en effet, la Chine, second partenaire commercial de l'Europe, n'est pas [un «concurrent»] **mais** [un «partenaire stratégique» qu'il convient de ne plus humilier].   (Mult-Ed-FR – Libération)

(81)   For the first time since March 2012, the Conservatives are in first place, **though** more because of a Labour slide of six points since April than because of a Tory surge   (Mult-Ed-EN – The Guardian).

(82)   La couverture de Charlie ? Un modèle d'intelligence politique. Beaucoup attendaient une provocation, d'autres craignaient un recul. Rien de tout cela. Le prophète Mahomet est au rendez-vous, **mais** dans un rôle positif, avec un soupçon de tendresse.   (Mult-Ed-FR – Libération)

iii. Thirdly, for a given marker to be weeded out as phrasal, the second segment in the relation must contain one single functional element – as in (83) and (84), where S2 only consists of a subject complement. By contrast, as stated by Halliday and Matthiessen (2004: 487), "as soon as more than one element is involved, we have to analyse the complexing at clause rank and posit [verbal] ellipsis in one of the clauses". Therefore, the coordinators in both (85) and (86) are included within the set of conjunctive markers, as the segments that they introduce contain two distinct functional elements, viz. a temporal adjunct and a subject complement in (85); and a subject together with its complement in (86).

(83)   But the inquiry has to be thorough, and our real criticism of the Home Secretary's appointment of Ms Woolf is not her choice of person, **but** her failure to provide the inquiry's terms of reference.

(Mult-Ed-EN – The Independent)

(84)   L'indépendance n'est pas un concept abstrait, **mais** une boussole.

(Mult-Ed-FR – Le Monde)

(85)   The case for greater flexibility is easily advanced, **but** at this juncture much more difficult to attain.   (Mult-Ed-EN – The Telegraph)

(86)   La durée effective du travail et le revenu n'auront pas bougé, l'effet sur l'emploi sera vraisemblablement nul, **mais** le coût fiscal à terme potentiellement très important.   (Mult-Ed-FR – Libération)

iv.   Finally, for a given segment to be considered phrasal there may be no 'strong' punctuation mark between S1 and S2. Rather, when a verbless segment is isolated from the surrounding discourse by means of strong punctuation (viz. typically, by full stops), as in (87) and (88), it is considered to function as a minor clause (see above for a definition). Thus, although the markers in examples such as (87) and (88) below both meet the three criteria for phrasal status listed above, they are nevertheless included within the set of conjunctive markers analysed in the present study. In addition, it must also be noted that examples such as (89) and (90), where neither of the segments related by the marker contains a verb phrase, are also included in the data set: rather than being analysed as phrasal, such markers are considered to link two minor clauses together.

(87)   Voir il y a quelques jours une petite fille accueillir la garde des Sceaux aux cris de «la guenon, mange ta banane» n'est pas quelque chose d'anodin. **Plutôt** une alerte sur l'état de la société et sur ses dérives.

(Mult-Ed-FR – Libération)

(88)   Refaire ses mérites ou ses méfaits, à l'endroit ou à l'envers, c'est certes fort passionnant. **Mais** pas vraiment d'avenir.   (Mult-Ed-FR – Le Monde)

(89)   That will further discourage Labour from saying anything pro-European before the general election and could embolden British Eurosceptics, and the Eurosceptic press, to press for a more radical agenda of disengagement. Not necessarily a body blow to Labour, **but** a severe constraint **nevertheless**
(Mult-Ed-EN – The Guardian).

(90)   Quant à Nafissatou Diallo, elle n'est peut-être pas celle que l'on croyait. Pas une plaignante au-dessus de tout soupçon, **mais** une femme intéressée, affabulatrice, aux relations improbables.   (Mult-Ed-FR – Le Figaro)

Based on the four criteria described above, I put forward a three-step decision process which was applied to all the occurrences of the markers encountered in the corpus data, making it possible to determine with as much certainty as possible whether they could be considered as true conjunctive markers with respect to the nature of the segments that they linked. This decision process is represented in Figure 12 below. Note that the procedure mostly focuses on the features of S2 in particular. The reason is that, because of their propensity to involve verbal ellipsis,

the status of these segments was generally much harder to grasp than that of S1 – which was usually much more straightforward.
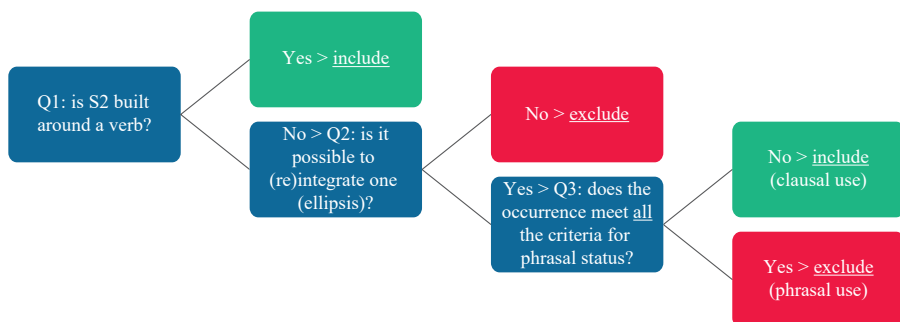
Figure 12. Decision tree for discourse segmentation

As this figure shows, for each potential CM found in the corpus, the first question asked was whether S2 was built around a verb. If it was the case, S2 was considered as clausal, and the CM token was included in the data set. If the answer to this first question was negative, I moved on to the next question in the decision tree, viz. is it possible to restore a verb in S2? This stage was meant to identify the instances in which S2 involved verbal ellipsis. If, as is the case in Examples (91) and (92), it was impossible to restore a verb in S2 (sometimes simply because S1 itself was not verbal), the occurrence was considered as phrasal, and weeded out from the data set. If, on the other hand, it was possible to reinsert a verb in S2 – as is the case in (93) and (94), where the elliptical segments are provided between square brackets – I applied the four criteria for phrasal status described above, so as to determine whether S2 was to be considered as a true verbless clause, or merely corresponded to a phrasal segment. Occurrences which met *all four* criteria for phrasal status were removed from the data set, whereas the tokens which failed to meet *at least one* of these four criteria were considered as instances of conjunctive markers introducing a verbless segment.

(91)   By Monday, in his tone, **though** sadly not in his specific proposals, Mr Cameron had calmed down.   (Mult-Ed-EN – The Guardian)

(92)   Renzi dit s'en accommoder, porté par des sondages favorables **bien qu'**en baisse.   (Mult-Ed-FR – Libération)

(93)   Unemployment and inflation are low; interest rates may be rising **but** [they are rising] from a very low base.   (Mult-Ed-EN – The Observer)

(94)   L'Enfant Bleu a perdu, **mais** [il a perdu] dans des conditions qui lui laissent augurer une prochaine victoire.   (Mult-Ed-FR – Le Figaro)

This segmentation procedure made it possible to identify and discard a very large majority of the markers used below the level of the clause. There remained, however, one specific set of cases which could not be adequately handled by means of the three-step decision process described above: in French, a number of conjunctive adjuncts – most of which were instances of the marker *pourant* – were used between (i) a noun phrase and (ii) another noun phrase or an adjective phrase which modified it, as in Examples (95) to (97).

(95)   Certains, **pourtant** peu suspects d'euroscepticisme, ont pu écrire que «les Français se sentaient comme des étrangers dans leur propre pays».

(Mult-Ed-FR – Le Figaro)

(96)   Pour France-Espagne, François Hollande, **pourtant** grand amateur de football, avait prédit un match nul.   (Mult-Ed-FR – Libération)

(97)   Pour le référendum sur Maastricht, M. Chirac, **pourtant** très eurosceptique, avait bien compris qu'en vue de sa nouvelle candidature à l'élection présidentielle de 1995, il n'avait pas d'autre choix que le "oui".

(Mult-Ed-FR – Le Monde)

Strictly speaking, these instances ought to be discarded from the data set, as they fail to fulfil the condition required by the second step of the decision process represented in Figure 12: in none of the examples above is it possible to reinsert a verb in the segment introduced by the CM (unless a relative pronoun is also added, e.g. *Mr Chirac, qui est pourtant très eurosceptique…*). Yet, when comparing such uses of *pourtant* with some instances of subordinators in both French and, especially English, it was felt somewhat problematic to weed them out altogether. If we compare the examples above with the verbless uses of the subordinators in (98) to (101), these borderline occurrences of French *pourtant* appear to be strikingly similar to some subordinator uses kept in the data set: in fact, it would be perfectly possible to replace each of the occurrences of *pourtant* above by subordinators such as *bien que* or *quoique*. Interestingly, these examples may thus shed light on the potentiality of some French CAs – and of *pourtant* in particular – to be used in ways that are normally more typical of subordinators (see Chapter 7 for more on this).

(98)   But the tanks, **although** obviously an instrument of Russian policy, are less important than the informational dominance which the regime has established.   (Mult-Ed-EN – The Guardian)

(99)   That is why the directness of Ukip, **even if** at times a little unsophisticated, struck a chord with voters.   (Mult-Ed-EN – The Telegraph)

(100) C'est dans ce contexte qu'on doit interpréter le geste du secrétaire général de l'enseignement catholique. Sa lettre, **quoique** anodine dans sa rédaction, s'inscrit évidemment dans l'effort général de mobilisation du monde catholique contre le gouvernement Ayrault.

(Mult-Ed-FR – Le Nouvel Observateur)

(101)   Pourtant, ce problème d'insertion, **bien que** très visible, est sans commune
        mesure avec les problèmes que doivent affronter leur vie durant la très
        grande majorité de ces Européens oubliés du progress.

                                                        (Mult-Ed-FR – Le Monde)

In view of the apparent structural resemblance between the uses of *pourtant* exemplified above, and the verbless uses of subordinators, it was felt to be rather inconsistent to discard the former instances from the data set, while keeping the latter.
One undesirable consequence of such a decision would have been to overlook
some typically-French ways of expressing relationships of contrast between two
discourse units. These examples were therefore kept in the data set. However, as
they did not strictly meet the required conditions to qualify as markers used at the
clausal level, these CM tokens were categorised as 'phrasal-clausal'. Note that in
any case, the markers labelled 'phrasal-clausal' only represent a minor portion of
markers in the corpus: they account for 87 instances, viz. a mere 0.4% of the total
number of CM tokens included in the data set.

6.2.4   Coding the syntactic features of the host clause

In the previous sections, I presented a number of criteria used to clearly determine
(i) which occurrences of the markers extracted from the corpus data could be considered to perform a truly conjunctive function by linking units at the clausal level
(or above), and (ii) which instances rather operated at the phrasal level, and thus
should be weeded out from the data set. Making the criteria used to distinguish
clausal from non-clausal segments fully explicit is essential in a study on conjunctive markers, as it may have profound consequences on the corpus results. For
example, if the present study had considered neither the possibility for clauses to
be verbless, nor the potential of embedded clauses to function as discourse segments – as was done in Carlson and Marcu's (2001) study, for example – as much
as 5% (viz. 797 CM instances) of the final data set obtained here would have been
disregarded – corresponding to 4.5% of the English, and nearly 6% of the French
CM tokens.

Alongside its key importance to ensure both the validity of the disambiguation and the replicability of the corpus analysis, the segmentation process also
shed light on the great variety of syntactic structures in which conjunctive markers
of contrast may be involved, while also suggesting that these syntactic patterns are
likely to differ not only between languages (e.g. with French displaying a larger
proportion of CMs used in verbless clauses than English), but also according
to the lexical item investigated (see e.g. the phrasal-clausal uses associated with
the French CA *pourtant*, as discussed above). This observation prompted me to
operationalise the reflections made on discourse segmentation by systematically
coding the syntactic patterns of the conjunctive markers in my data set – rather

than resorting to syntax simply to disambiguate the corpus data. For reasons of feasibility, I decided to focus on the features of the clauses hosting the CMs, i.e. I concentrated on S2. For each CM occurrence in the data set, the features of the host clause were coded along the following dimensions:

i.   CM CATEGORY, viz. coordinator, subordinator or conjunctive adjunct;
ii.   STRUCTURAL TYPE of the clause hosting the CM, viz. finite, non-finite or verb-less clause;
iii. RANK STATUS of the host clause, viz. main clause, hypotactic clause, embed-ded clause, or minor clause.

The multilayer coding design makes it possible for each level of these three vari-ables to combine with any level of the other variables, as illustrated in Table 7. Examples of conjunctive markers used in each type of syntactic structure are pro-vided in Table 8. Markers from each category are exemplified in each of the syntac-tic classes – at least whenever it was possible to find examples for all three types of CMs. The results of the syntactic coding are presented in Chapter 7.

Table 7. Example of multilayer syntactic coding of the corpus data: the marker *though*

CONCORDANCE	CM CATEGORY	CLAUSE TYPE	RANK STATUS	TEXT ID
The benefits, **though**, would include fewer of the sordid stories of alleged corruption	Conjunctive adjunct	Finite	Main	\<E-GU-v1-493\>
What, **though**, of the choice itself?	Conjunctive adjunct	Verbless	Minor	\<E-GU-190914-2\>
Though condemnation has been wide, multiple factors have worked in his favour.	Subordinator	Finite	Hypotactic	\<E-GU-140514-2\>
Mr Blair used it, **though** without giving it prime importance, in his attempts to convince the British public of the rightness of his decisions.	Subordinator	Non-finite	Hypotactic	\<E-DT-v1-501\>
Though not an uncommon attitude, it is arrogant and even absurd.	Subordinator	Verbless	Hypotactic	\<E-DT-v1-413\>
Lucky escape I'd say. **Though** not for BB.	Subordinator	Verbless	Minor	\<E-FT-010814-3\>

With respect to the syntactic coding of the conjunctive markers, it is useful to spec-ify how cases of syntactic 'nesting' were handled in the corpus data. 'Nesting' refers

Table 8. Syntactic patterns of use of conjunctive markers of contrast

SYNTACTIC CATEGORY		ENGLISH EXAMPLES	FRENCH EXAMPLES
CLAUSE TYPE	FINITE	– Merkel was not keen on Juncker either, **but** unlike Cameron, she <u>emerges</u> unscathed from a bad situation (Mult-Ed-EN – The Observer). – **Although** the Brics nations now <u>account</u> for about 20 per cent of global output, they get only 10.3 per cent of the votes at the IMF (Mult-Ed-EN – The Financial Times). – They are critical of M Aristide's rule […]. **However**, they <u>recognise</u> that he was duly elected four years ago and that his immediate departure could create a power vacuum inviting even greater chaos than at present (Mult-Ed-EN – The Telegraph).	– La loi est un corset **mais** elle <u>garantit</u> souvent les faibles (Mult-Ed-FR – Libération). – **Bien que** l'économie ne <u>soit</u> pas le fort de chacun, on a pu lire, comprendre et admirer l'entretien publié dans nos Débats de «l'Obs.» de la semaine dernière (Mult-Ed-FR – Le Nouvel Observateur). – L'idée que les Français les moins favorisés vont naturellement vers la gauche est malheureusement réfutée depuis longtemps. **Au contraire**, la gauche <u>fait</u> ses meilleurs scores au centre des grandes villes (Mult-Ed-FR – Le Nouvel Observateur).
	NON-FINITE	– European regulators of the single market should strive not to "harmonise" taxes and social policy across the continent, **but** <u>to foster</u> growth on the principle of comparative advantage among differentiated economies (Mult-Ed-EN – The Telegraph). – **Though** <u>commissioned</u> by a public-school dominated government, Alan Milburn has done a decent job of setting out some uncomfortable facts about the narrow background from which Britain's great and not-so-good are drawn (Mult-Ed-EN – The Guardian). – Ed Miliband has accused David Cameron of being a cheerleader for the deal, <u>arguing</u> **instead** for jealously protecting AstraZeneca's independence (Mult-Ed-EN – The Telegraph).	– Le chef de l'Etat a certes dit la nécessité «d'aller plus loin», **mais** sans <u>donner</u> les clés de cet «ailleurs» (Mult-Ed-FR – Libération). – Israël, **bien qu'**<u>ayant</u> implicitement accepté les termes de cette résolution par le fait même de son admission à l'ONU, ne les a jamais reconnus de facto (Mult-Ed-FR – Libération). – Si une ONG dérape, pourquoi les autorités françaises et surtout le Quai d'Orsay, <u>dirigé</u> **pourtant** par un ancien humanitaire, ont-ils finalement laissé faire (Mult-Ed-FR – Libération).

(*Continued*)

Table 8. (*Continued*)

Syntactic category		English examples	French examples
	Verbless	– Should Austen, Dickens, Hardy and Shakespeare be part of such a list? Of course they should. **But** adjacent to Zadie Smith, Sarah Waters and Meera Syal (Mult-Ed-EN – The Observer). – Dr David Kelly conveyed, **albeit** in a somewhat mangled and self-serving form, the concern of some members of the intelligence world about that misrepresentation to the BBC reporter Andrew Gilligan (Mult-Ed-EN – The Telegraph). – We in Britain, and especially Northern Ireland, are reasonably familiar with the concept, having endured the smaller scale bombing campaign of the IRA in the two decades before 1996. Not as familiar as the Israelis, **however** […] (Mult-Ed-EN – The Independent).	– Les Etats-Unis ont-ils terminé leur règne d'hyper puissance ? Peut-être, **mais** pas celui de super (Mult-Ed-FR – Le Nouvel Observateur). – **Quoique** riche, il voyait l'entreprise, non comme une simple machine à fric, mais comme une institution de la société (Mult-Ed-FR – Libération). – Que se passe-t-il donc ? Oh ! rien d'excessivement drolatique en fait. Un constat dramatique **plutôt** (Mult-Ed-FR – Le Monde).
Rank status	Main	– There has never been more data in the public arena than now, **but** there has never been less shared confidence in it (Mult-Ed-EN – The Guardian). – The EU represents an inward-looking attempt to preserve the wealth of a continent declining in economic power. The Commonwealth, **by contrast**, is outward-looking […] (Mult-Ed-EN – The Telegraph).	– Le film n'est pas qu'un renvoi historique à la Judée, **mais** il se présente comme représentation compassionnelle de la propre situation des Etats-Unis en tant que victimes d'une violence injuste, celle du terrorisme international (Mult-Ed-FR – Libération). – Nul ne peut contester l'ampleur du propos, tout comme la gravité du ton démontrait la conscience des dangers. Les questions restent **cependant** ouvertes (Mult-Ed-Fr – Le Figaro).

HYPOTACTIC	– If it chooses to use the pound **but** rejects close fiscal and monetary union with England, an independent Scotland would be floating itself on stormy seas indeed (Mult-Ed-EN – The Independent). – **While** that was a welcome decision, the WHO regrettably failed to combine it with advice on who should receive any available medicines (Mult-Ed-EN – The Financial Times). – Should he prove refractory, and slip back into bad behaviour, conservatives will be vindicated. If, **however**, he now behaves in exemplary fashion, the liberals will be encouraged in their contention that improving individual circumstances is the route to a safer society (Mult-Ed-EN – The Times).	– Le chef de l'Etat s'est limité, pour le moment, à souhaiter des «aménagements légaux» de ce qu'il considère encore comme «un droit acquis», **mais** sans fixer précisément la philosophie de la refonte (Mult-Ed-FR – Le Monde). – [S]'il ne fait pas de doute que pour l'enfant rien ne vaut la cellule naturelle du couple hétérosexuel stable, celui-ci est loin de demeurer la norme (Mult-Ed-FR – Libération). – Nous avons étonnamment peu de réponses à ces questions et à tant d'autres, qui sont **pourtant** essentielles à une allocation intelligente des ressources et à une aide au développement plus efficace (Mult-Ed-FR – Libération).
EMBEDDED	– What can be done to tackle extremist Islamist groups who are not violent **but** whose ideologies are inflammatory and might trigger violence by some individuals? (Mult-Ed-EN – The Financial Times). – As such, he reflects British public opinion, which is torn between the fear that nothing in this arena seems to work and the instinct that **nevertheless** something must be done (Mult-Ed-EN – The Guardian).	– Notre volonté n'est pas de jouer le peuple contre l'élite **mais** que chacun, supposé intellectuel ou non, parisien ou provincial, fonctionnaire ou salarié, participe à élaborer le projet collectif (Mult-Ed-FR – Libération). – Mais le petit compromis arraché de haute lutte dans une atmosphère délétère par la Maison Blanche a surtout mis en lumière la faiblesse du pouvoir politique dans un pays qui demeure **pourtant** la première puissance mondiale (Mult-Ed-FR – Libération).

(*Continued*)

Table 8. (*Continued*)

SYNTACTIC CATEGORY	ENGLISH EXAMPLES	FRENCH EXAMPLES
MINOR	– Humiliation for some. **But** for those who may never win academic glory, a chance to shine (Mult-Ed-EN – The Guardian). – But the hope is that it will enable Mr Cameron to distinguish himself from both Mr Farage and Ed Miliband […] as the man with a proven track record of delivering big, difficult but important things. **Even if** it takes some time, and hard work, to get there (Mult-Ed-EN – The Telegraph). – Indubitably […] the US has made great strides in the past half-century towards achieving racial equality. Less so, **however**, in the field of policing (Mult-Ed-EN – The Independent).	– George Bush et Vladimir Poutine, alliés de circonstance contre un même ennemi, jouent une partition commune. **Mais** pas dans le même but (Mult-Ed-FR – Le Figaro). – Ce jeune homme de 27 ans n'est pas le seul Français à s'être engagé dans le djihad. **Même si** les autres, comme Hervé Loiseau, retrouvé mort sur les contreforts de l'Hindu Kuch, ont souvent des origines arabes (Mult-Ed-FR – Le Figaro). – Le moment est admirablement choisi pour nous servir cette recette, la révision constitutionnelle annoncée paraît novatrice et audacieuse, et cette réforme peut forger une nouvelle architecture des pouvoirs. A deux conditions, **cependant** (Mult-Ed-FR – Le Figaro).

to the phenomenon whereby a dependent clause may be attached to a clause which is itself dependent on another clause. In Example (102), for instance, a reporting *that*-clause is hypotactically dependent on a main clause, and itself contains an embedded clause modifying the noun *ways* that functions as the head of its complement (see e.g. Halliday & Matthiessen 2004: 376; Eggins 2004: 285; Bloor & Bloor 2004: 170 for more on nesting). When confronted with such examples, the syntactic status of the clause hosting the CM was always coded with respect to its closest, most direct level of syntactic integration. Thus, for example, the CA *nevertheless* in Example (103) was considered to occur in an embedded clause, even though the embedded clause containing the marker was itself included within a hypotactic (viz. non-defining relative) clause. Conversely, in (104), the CM *although* is considered to be included in a hypotactic clause, despite the fact that this hypotactic clause is itself part of an embedded (viz. defining relative) clause modifying the noun *claim*.

(102) No one disputes [that there are various ways [in which the content of an utterance can go beyond sentence meaning]].

(adapted from Bloor & Bloor 2004: 170)

(103) As such, he reflects British public opinion, [which is torn between the fear [that nothing in this arena seems to work] and the instinct [that **nevertheless** something must be done]].   (Mult-Ed-EN – The Guardian)

(104) In particular, the claim [that they will help protect us from a 11 September-style terrorist attack, [**although** the cards were envisaged before then]], just seems cynical.   (Mult-Ed-EN – The Observer)

6.3   Conclusion

This chapter has described the process of manual disambiguation of the corpus data, which consisted of two main stages, viz. semantic disambiguation and syntactic segmentation. In the first part of the chapter, emphasis was laid on the high degree of polysemy of some of the markers investigated. In addition to conveying a logico-semantic relation of contrast, some of the CMs in the list used as a basis for the corpus study were also shown to express a range of other meanings, such as temporality, addition or emphasis. Consequently, the set of CM occurrences that had been extracted automatically from the corpus had to be disambiguated manually in context to only keep the occurrences of these markers which conveyed a contrastive meaning. This was far from an easy task, as a large number of CM tokens appeared to exhibit highly ambiguous meanings, which genuinely straddled the boundary between semantic categories. Rather than try to smooth over these ambiguities and attempt to force each polyfunctional token into a well-defined category, I made use of double semantic tags such as 'temporal-contrastive'

or 'additive-contrastive'. All the CM tokens which were found to combine contrast with another meaning were kept for subsequent analysis. The decision to allow for double tagging was found to most faithfully reflect the linguistic reality emerging from the corpus data, while also preventing arbitrary decisions from being made.

The second part of the chapter was dedicated to syntactic segmentation, viz. the process aiming to identify the markers which signalled a relationship between two clauses or larger units, while weeding out those which were used to link segments at phrasal level – and thus were not considered to perform a discourse function. Special care was taken to provide a definition of the clause that would be equally appropriate for the description of English and French CM usage. More precisely, the comparison of the English and the French literature made clear the necessity to go beyond a fully verb-based definition of the clause, in order to account for the seemingly greater tolerance of French for clauses lacking a verb phrase. Consequently, the definition of the clause adopted in the present study relied on the presence of a predicate rather than a verb. However, while a verb-based definition of the clause would have been too restrictive to allow for a reliable account of English and French CM usage, the predicate-based definition turned out to be rather too inclusive: it did not make it possible to systematically differentiate between clausal and phrasal uses of the markers in the corpus. The solution to that problem was to complement this broad definition of the clause with a clear characterisation of the set of items which were considered to be unquestionably phrasal. The segmentation of the corpus data also shed light on the large diversity of syntactic structures in which English and French conjunctive markers of contrast may be included, prompting me to code the syntactic features of the clauses hosting the markers in my data set. The results of these analyses are presented in the following chapter.

All in all, this chapter provided a very transparent description of the criteria used to select my data set, both from a semantic and a syntactic point of view. Such concerns of transparency are crucial in a study dealing with such complex and polyfunctional linguistic items as conjunctive markers. Given the functional 'plasticity' of these items (which may be used to express a diversity of meanings, and relate segments of various sizes), the absence of clearly circumscribed criteria of definition and selection of the markers included in a given study severely impedes both the replicability of the analyses performed, and their comparability with the results obtained in other studies. The semantic criteria used for disambiguating and coding conjunctive markers are usually granted quite a lot of attention in discourse research. For example, a large number of studies in recent years have been working towards a greater standardisation of the semantic frameworks of annotation of discourse relations, both within and across languages (see e.g. Benamara & Taboada 2015; Demberg et al. 2017; Sanders et al. 2018; and, more

generally, most of the research carried out in the framework of the Textlink COST action, led by Prof. L. Degand between 2014 and 2018).[5] By contrast, as was made clear at several points in this chapter, far less attention has been devoted to issues of segmentation in the discourse research to date. Many researchers simply state that markers are considered to perform a discourse function when they are used to relate clauses or larger units, without clearly specifying how they define the clause. Yet, as I hope to have demonstrated here, the clause is far from a straightforward concept, especially when approached from a contrastive perspective. As a result, a lack of explicitness with respect to the definition of the clause is likely to lead to discrepancies in the types of units viewed as basic discourse segments across different studies, which can in turn severely affect the comparability of their findings (see also Sanders et al. 2018: 54–55 for a similar observation). One important contribution of this chapter has been to delve deeper into questions of discourse segmentation for the cross-linguistic study of conjunctive markers, by raising a number of important issues that may arise when identifying explicit markers of discourse relations in authentic, multilingual corpus data, and putting forward very clear syntactic criteria of identification of the markers truly performing a conjunctive function.

In total, the form-based automatic extraction of the markers from the corpus gave rise to a set of 26,013 hits in the Mult-Ed subcorpus (where all three categories of markers were analysed), and 13,881 hits in the LOCRA subcorpus (where only the conjunctive adjuncts were investigated). After this data had been disambiguated both semantically and syntactically, I ended up with a total of 15,364 conjunctive markers of contrast in Mult-Ed, and 8,635 conjunctive adjuncts of contrast in LOCRA. This means that about 40% of the set of tokens which were extracted automatically from the corpus had to be weeded out manually. These figures make clear the extensive amount of manual work required by discourse analysis. After a very 'quick and easy' automatic extraction of the markers in my list from the corpus, I had to go through a very slow and painstaking process of manual selection of the data – which in fact constituted by far the most time-consuming stage of the present study. This means that, prior to even starting to analyse the data quantitatively in relation to the research questions asked in this book, it was necessary to carry out a considerable amount of essentially 'qualitative' work.

---

5. <https://www.cost.eu/actions/IS1312> (12 April 2020).

Frequency and patterns of use of English and French conjunctive markers of contrast

7.1 Introduction

The present chapter aims to provide a broad contrastive overview of the use of conjunctive markers of contrast in English and French written language. It revolves around three main research questions:

i.   Do English and French display differences in their frequency of use of explicit conjunctive markers of contrast?
ii. Do English and French differ in the types of conjunctive markers (i.e. conjunctive adjuncts, subordinators or coordinators) that they tend to prefer to signal discourse relations of contrast?
iii. In what types of syntactic structures are English and French conjunctive markers of contrast involved? Do English and French differ in that respect?

Based on the English-French contrastive literature, I was able to formulate hypotheses for two of these three research questions. Firstly, with respect to the overall frequency of use of explicit CMs of contrast, it can be hypothesised that French will display a higher frequency of CMs than English. As explained in Chapter 3, the dominant claim in the literature is that French tends to signal logico-semantic relationships by means of explicit CMs more frequently than English. English, on the other hand, is argued to more often rely on the implicit juxtaposition of discourse units, thereby leaving it to the reader to infer the connections between them (see e.g. Vinay & Darbelnet 1995: 234; Hervey & Higgins 1992: 49). Secondly, regarding the types of markers preferred by each language to signal relationships between two discourse segments, a number of researchers have argued that relationships between clauses are commonly expressed by means of coordination in English, and subordination in French (see e.g. Guillemin-Flescher 1981: 143; Chuquet & Paillard 1987: 151; Delisle 2013: 601). In addition, conjunctive adjuncts have been claimed to be more common in French than in English (Vinay & Darbelnet 1995: 238; Chuquet & Paillard 2017: 40; see Chapter 3

for more details). As a result, it can be expected that logico-semantic relations of contrast will tend to be signalled more frequently by means of coordinators in English, and subordinators as well as CAs in French.

Concerning the types of syntactic structures in which English and French CMs of contrast are included, by contrast, no clear hypothesis can be formulated at this stage. The reason is that, although the English-French contrastive literature contains a series of claims pertaining to the types of syntactic structures which are particularly common in one language as compared to the other, to this day there is very little consensus on these issues. As a matter of fact, it is not uncommon to find conflicting evidence within a single study. Cosme (2008b), for instance, carried out a translation-based analysis aiming to test the hypothesis that French tends to express processes in nominal (or phrasal) form more commonly than English. In her corpus, she found not only (i) patterns of translational correspondence between a French phrasal constituent and an English clausal constituent, but also (ii) reverse patterns of equivalence, where English phrasal constituents give rise to clausal constituents in French. She explains that both patterns are equally frequent in her translation corpus. Another good example is Astington's (1983) book devoted to the translational patterns of equivalence between English and French. In the first part of his book, which is dedicated to the translation problems encountered when working in the French-English translation direction, Astington (ibid.: 45) shows that the translation of some French nominal groups into English requires the addition of a verbal element, as the "increasing tendency towards nominalization in contemporary French" cannot be transferred directly into English. This may lead us to believe that verbless clauses will be more common in French than in English. However, in the second part of his book, devoted to the English-French translation direction, Astington (ibid.: 95) discusses a number of cases where English nouns need to be translated by a French verb. Thus, there is no sufficient agreement in the contrastive literature to allow for the formulation of a reasonable hypothesis concerning the syntactic patterning of English and French CMs of contrast.

The corpus analyses presented in this chapter are based primarily on the Mult-Ed comparable corpus of quality newspaper editorials. Some of the findings obtained on the basis of Mult-Ed are complemented with smaller-scale analyses of the LOCRA comparable corpus of research articles, so as to integrate register-sensitive considerations into some of the comparisons. All the frequencies are provided as relative frequencies per million words. The differences uncovered between the two languages (and, when relevant, between the registers) are tested for statistical significance using chi-square tests of independence. Finally, CMs that did not reach the threshold of five occurrences per million words were excluded from the analysis to avoid distracting the reader with the discussion of markers which play

an extremely limited role in the English and French systems of contrastive linking. The decision to exclude the very infrequent markers has no bearing whatsoever on the tendencies and differences which are outlined in this chapter since, in both languages, the discarded markers together represent well below one percent of the total number of occurrences of CMs identified in the corpus.

The present chapter is divided into three main parts. Section 7.2 compares the frequency of conjunctive markers in English and French, both overall and per CM category, thus answering the first two research questions. In addition, it also includes more detailed descriptions of the frequencies of individual markers of contrast identified in each subcorpus. Section 7.3 focuses on the syntactic patterning of CMs of contrast in English and French, comparing the types of syntactic structures in which these items tend to be included in each language, and discussing these differences in the light of both discourse and lexical factors. Finally, Section 7.4 complements the analyses carried out in the previous sections by adopting a register-sensitive approach to the contrastive study of CMs.

7.2 Frequencies of conjunctive markers of contrast in English and French editorials

7.2.1 Overall frequency of conjunctive markers of contrast

The objective of this section is to test the widespread claim that French tends to use more explicit conjunctive markers than English, focusing on the logico-semantic relation of contrast. The overall relative frequencies of CMs of contrast in the Mult-Ed corpus of editorials are presented in Table 9. The standardised residuals emerging from the chi-square comparison of these frequencies are provided between brackets next to the relative frequencies.

Table 9. Relative frequency of conjunctive markers of contrast in English and French (per million words)

	ENGLISH	FRENCH
REL. FREQUENCY OF CMs OF CONTRAST	8,656 (+26.0)	5,629 (−25.4)

As shown in Table 9, the relative frequency of CMs of contrast is markedly (viz. 1.54 times) higher in the English than in the French subcorpus. Therefore, not only is the hypothesis that French tends to be more explicitly cohesive than English strongly rejected by the corpus results, but the opposite is in fact true: it is English that appears to display a higher frequency of explicit CMs than French,

at least for the meaning relation of contrast. The application of a chi-square test of independence on the data reveals that the difference between English and French is statistically significant (χ^2 = 1331.7; df = 1; p < 0.001; φ = 0.02, with significant standardised residuals for both values).[1]

In the face of these very surprising results, it is worth delving into some of the possible reasons that may account for such a strong rejection of the initial hypothesis. A first question that arises is whether the results obtained here are specific to the logico-semantic relation of contrast. As highlighted in Chapter 3, the claims pertaining to cross-linguistic differences in degree of cohesive explicitness are generally presented as being valid for the English-French language pair as a whole. Yet, some studies have suggested that cross-linguistic differences in frequencies of CMs may vary in function of the type of logico-semantic relation investigated. In her study of Spanish and English explicit markers of premise-conclusion in research articles, for instance, Moreno (1998: 550) questions the relevance of general statements concerning cross-linguistic differences in use of CMs:

> To what extent, for example, might it be of interest to know that English is more explicit than Spanish in the field of connectives? The question rapidly arises as to whether the same would be true for all the coherence relations susceptible to being expressed by a connective.

In a similar vein, Cuenca and Bach (2007) uncover a higher frequency of explicit markers of reformulation in Spanish than in English academic prose, which contrasts with previous findings on CM use in these two languages. They assert that this particular difference is likely specific to the reformulation relation and testifies to more digressive and indirect strategies of textual development in Spanish as compared to English (ibid.: 155; see also Chapter 3 for more details). In that context, it may be that French does indeed have a greater tendency towards cohesive explicitness than English for some relations, but that for other types of meanings – including contrastive ones – English requires more explicit relational

---

1.   As a reminder, the φ values report the effect size associated with each chi-square test. Note that, according to Cohen's scale, the effect size obtained here is really fairly small – as are most of the φ values reported in the discussions that follow – suggesting that the variable under study (viz. language) only has a limited effect on the frequencies observed in the corpus data. However, as already underlined in Chapter 5, some researchers have recently advocated interpreting effect sizes with respect to the values typically reported in the field in which they are obtained. Yet, to my knowledge, no study in corpus-based contrastive linguistics has yet systematically reported the effect sizes associated with their chi-square results. Consequently, at this stage, I do not have sufficient information at my disposal to truly interpret my effect sizes, and I mainly provide them so that they can serve as a basis for interpretation and comparison in future corpus-based contrastive research (see Chapter 5 for more details).

markers than French. Such a hypothesis is in line with some previous intuition-based research on English and French CM use. Whereas the dominant claim in the literature is that French uses more CMs than English, a few studies have also made the opposite claim, i.e. they have argued that it is English that tends to require more explicit marking than French (see Chapter 3). Interestingly, among these rare studies, three focus specifically on relations of contrast. Based on a few literary examples in French, for example, Gallagher (1995) observes that contrastive relations are frequently left unmarked in contemporary written French. He adds that in translation into both English and German, explicit contrastive links generally need to be restored, failing which the coherence of the target text will be seriously hampered (ibid: 231):

> In English as well as German, the deletion of adversative and concessive connectives is infrequent. Where French tends to resort to juxtaposition or punctuation, the French-English translator will thus have to clarify and disambiguate intra- and inter-sentential relationships by using adversative and concessive connectives [my translation].

Likewise, Mason (1998: 172) notes that it is frequent for French writers to omit markers of contrast in argumentative texts, whereas in English, such explicit markers are generally required for the text to be processed smoothly by the reader:

> [E]llipsis of [opposition markers] is sufficiently frequent in French for it to be a recognisable text strategy. The markedness of the structure constitutes no problem for the French reader. In translation into English, on the other hand, straightforward transfer of the ellipsis may create problems of coherence for target-text readers and consequently, […] there is a tendency for translators to restore explicit junction.

Finally, Guillemin-Flescher (1981: 189) highlights some cases in which two clauses that are simply juxtaposed in French need to be related by an explicit marker of contrast in translation into English. Therefore, whereas those who have argued that French tends to require fewer explicit signals of cohesion than English only form a minority in English-French contrastive linguistics, a fair portion of their claims seems to stem from the analysis of relations of contrast – a fact that is probably significant in view of the results obtained here. As these studies are small-scale, and rely on limited sets of hand-picked examples from translation, the corpus results presented in this chapter emerge as a valuable empirical confirmation of their observations, demonstrating that they may be generalised beyond the small data sets on which they are based.

Other research provides more indirect support in favour of the hypothesis that explicit CMs of contrast are particularly infrequent in French. Such evidence is found in Takagaki's (2011) study of CMs in French and Japanese. Based on a

bidirectional translation corpus in which she identifies cases of both CM additions and omissions, Takagaki aims to test the hypothesis that French is more explicitly cohesive than Japanese. She reports that, although her general hypothesis is confirmed, with French displaying a higher frequency of explicit CMs than Japanese overall, the two languages differ markedly in the types of logico-semantic relations which they tend to leave implicit. Whereas Japanese usually leaves temporal and causal relations unexpressed, in French adversative (i.e. contrastive) and additive relations are the ones which are usually left implicit. Another relevant study is the one by Granger and Tyson (1996), carried out within the framework of learner corpus research. Granger and Tyson set out to compare the use of explicit CMs in the writing of native speakers and French-speaking learners of English. Starting from the claims on CM usage found in the English-French contrastive literature, they hypothesise that French-speaking learners of English will tend to overuse CMs in their written productions, as a result of transfer from their mother tongue. Granger and Tyson's corpus results demonstrate that, while the hypothesis is confirmed for *some* categories of CMs (viz. additive, exemplifying or corroborating CMs), learners also tend to *underuse* some types of CMs, among which contrastive ones (ibid.: 20). The authors suggest that these patterns of over- and underuse may be traced back to diverging patterns of discourse organisation in the two languages, although they stress that such a hypothesis awaits to be tested empirically in a corpus-based contrastive study of English and French (ibid.). The results of both of these studies are in keeping with the hypothesis that the type of relation investigated here may have had a bearing on the results, with CMs of contrast being particularly 'uncommon' in French, at least as compared to markers of other logico-semantic relations.

However, this hypothesis remains speculative: in order to determine whether the results uncovered here are truly – or even partly – specific to contrast, it would be necessary to compare them to equivalent results obtained for CMs belonging to other logico-semantic categories. Unfortunately, this is not possible at this stage, since no comparable, onomasiological corpus-based comparisons of CM frequencies have yet been carried out on English and French written language. The only relevant quantitative results that I was able to find are those provided in Zufferey and Cartoni's (2012) study of equivalences between English and French backward causal CMs (e.g. *since, because, for*; *parce que, car, puisque*). Prior to investigating the translation of the twelve markers in their bidirectional translation corpus of parliamentary debates, they provide the relative frequency of each of these CMs in the original subparts of the corpus. Zufferey and Cartoni (ibid.: 239) report a frequency of 148 CMs per 100,000 words in English, against 164 CMs per 100,000 words in French. As opposed to the results presented in Table 9, these figures are thus in line with the claims commonly found in the contrastive literature, although the frequency difference is arguably limited. In addition, the

study by Crible (2018), which compares the use of discourse markers across a range of semantico-functional categories in a comparable corpus of English and French spoken language, also provides useful insights. Crible (2018: 82) reports that overall, spoken French displays a higher frequency of explicit markers than spoken English (60 vs 49 markers per thousand words). However, she also shows that the two languages differ in the types of discourse functions that they most frequently signal by means of discourse markers. For example, she shows that markers of consequence (e.g. *so, then; alors, donc*) are more frequent in English than in French, whereas markers of specification (e.g. *indeed, in fact*; *d'ailleurs, c'est-à-dire*) are more typical of French (ibid.: 107–108) – although the frequency differences that she reports are possibly too small to be statistically significant. The problem is that, as Crible's study is not restricted to conjunctive markers, but also includes a range of speech-specific, non-cohesive discourse markers (e.g. *well, you know, right; hein, écoute*), her results are not directly comparable to mine. Thus, it would be useful to replicate the present study for other logico-semantic relations, so as to know whether the results reported here are generalisable to all types of discourse relations, or specific to some of them. In any case, even though the present study does not allow for a categorical rejection – or confirmation – of the claims found in the English-French contrastive literature, it clearly demonstrates the sheer necessity of balancing these claims according to the type of semantic relation investigated.

In addition to contradicting the claims made in the previous contrastive literature, the fact that conjunctive markers of contrast, in particular, occur with such radically different frequencies in English and French also appears surprising in the light of some language-independent, cognitive principles of discourse organisation formulated in the literature. Although CMs in general have frequently been said to be optional linguistic units, (some) relations of contrast have sometimes been argued to be rather difficult to infer in the absence of an explicit signal, due to the relative complexity of the cognitive operations that they entail (see Chapter 2 for more details). Since CMs of contrast have been said to be relatively essential as compared to other types of CMs, we may have expected the frequency difference between the languages to be fairly limited. In view of the sizeable frequency difference uncovered between the two languages, it is therefore worth wondering whether French has alternative ways of expressing contrast in discourse, in addition to interclausal and intersentential CMs. As discussed in Chapter 2, CMs are not the only ways in which coherence relations may be signalled, and some researchers have underlined a number of alternative techniques – e.g. mood, modality, relative pronouns, etc. – which writers may resort to in order to express a given relation in discourse (e.g. Taboada 2006; Das & Taboada 2013). With respect to the expression of contrast, French does seem to have alternative resources which

can take on the function of conjunctive markers. Gallagher (1995: 217–218), for example, mentions emphatic pronouns such as *lui, eux,* etc.; introductory phrases such *quant à, pour [sa] part*; or disjuncts such as *malheureusement, au vrai, en fait*, etc. (see also Thompson & Zhou 2001 on the structuring functions of disjuncts). These strategies are exemplified in (1) to (3). Gallagher (ibid.: 205) explains that, unlike conjunctive markers in the strict sense (e.g. *cependant, toutefois*), which are often done without in original French, such devices are not easily removed without impeding the coherence of discourse.

(1)   Le règlement des lycées prescrivait les lectures à haute voix pendant les repas. Précaution contre l'agitation des esprits. Les oratoriens, **eux**, autorisaient les conversations au réfectoire.   (Gallagher 1995: 205)

(2)   Après le sermon, je ne sais quel individu ne se serait pas cru déjà brûlé, damné ; **quant à moi**, j'avais si froid aux jambes que je me suis aperçue ne pas être encore en enfer   (ibid.)

(3)   Pourquoi ai-je toujours passé mes vacances avec vous au lieu de voyager ? Je pourrais imaginer de belles raisons. **Au vrai**, il s'agissait pour moi de ne pas faire double dépense.   (ibid.: 216)

Another strategy that is commonly used to express a contrast between two French sentences is thematisation (see e.g. Rivelin-Constantin 1992; Poncharal 2005: 295; Rossette 2007: 40). Thematisation is typically achieved through the use of cleft and/or dislocated syntactic structures, as exemplified in (4) and (5), where the thematised structures implicitly convey a relation of opposition and correction, respectively.

(4)   <u>Bernard, il</u> drague Gigi. Et <u>toi, tu</u> couches avec Jean-Claude. Bref, tout le monde rigole, sauf moi.   (Rivelin-Constantin 1992: 164)

(5)   <u>Moi, j'</u>avais rien dit. Rien. <u>C'est</u> Arthur Ganate <u>qui</u> m'a fait parler. (ibid.: 186)

While most of these structures are also available in English, some of them tend to be much more marked, and therefore far less common than their French equivalents. Rivelin-Constantin (1992), for example, has shown that thematising structures are markedly more frequent in French than in English, and tend to disappear in translation from French into English (see also Trévise 1986). This is exemplified in (6), where the thematising signals are underlined in the French sentence, and disappear in the English translation.

(6)   FR. J'aurais bien voulu qu'il m'explique <u>celui-là</u>, pendant qu'il y était, <u>ce réserviste</u>, pourquoi j'avais pas de courage non plus <u>moi</u>, pour faire la guerre, comme tous les autres.

          EN. I'd have liked this reservist to explain, while he was about it, why I had no stomach [sic.] either to make war like everybody else.

          (Rivelin-Constantin 1992: 176)

Likewise, the use of emphatic pronouns is much more typical of French than English, which tends to leave most emphatic stresses implied in writing (see e.g. Vinay & Darbelnet 1995: 220–221; Hervey & Higgins 2002: 118). These alternative strategies available in French to express contrast may compensate – at least partly – for a comparatively low frequency of explicit conjunctive markers of the kind investigated in this study.

In addition, part of the difference in frequency of contrastive CMs between English and French may pertain to the fact that French has frequently been claimed to display a rather 'nominal style' of expression as compared to English. In other words, a number of contrastive linguists have stated that processes which tend to be expressed by means of a verb – i.e. in clausal form – in English are frequently expressed in nominal form – i.e. as a phrase – in French (see among others: van Hoof 1989: 62; James 1980: 116; Guillemin-Flescher 1981: 14–30; Hervey & Higgins 1992: 204; Vinay & Darbelnet 1995: 100; Salkoff 1999: 171). Chuquet and Paillard (1987: 20), for example, explain that "the tendency of English to actualise processes and insert them in a chronology frequently gives rise to the use of a verb phrase, where a noun phrase would be used in French" [my translation]. This tendency is exemplified in (7) and (8).

(7) FR. Mais, <u>par un effort de volonté</u>, ce spasme disparut.
    EN. <u>She made an effort of will</u> and the spasm passed.
                         (Guillemin-Flescher 1981: 15)

(8) FR. <u>À 35 ans</u>, il fait les marchés depuis 20 ans
    EN. <u>He is now 35</u> and has been doing the markers for the last 20 years.
                         (Chuquet & Paillard 1987: 22)

Against this backdrop, it can be postulated that French may tend to express a certain proportion of its relations of contrast at the phrasal level where, in similar contexts, English would resort to an interclausal or intersentential conjunctive marker – as is the case in (9), for instance. Such cases of contrastive linking, which belong to the realm of logical metaphor (see Chapter 2 for a definition) were not included in the analysis, which only looked at markers expressing a relationship between two clauses or larger units.

(9) FR. Malgré mon absence de Paris…
    EN. Although I was away from Paris….   (Hervey & Higgins 1992: 211)

The more nominal nature of French may also entail a greater propensity towards other forms of logical metaphor, where, for example, the logical relationship is expressed through the use of a verb (instead of a CM), while the two related segments are nominalised (instead of being expressed as clauses), as in Example (10), extracted from the Mult-Ed corpus.

(10)   La sévérité de l'Allemagne **contraste** avec sa retenue sur le retour de
         Vladimir Poutine au Kremlin.   (Mult-Ed-FR – Le Monde)

In a future study, it would be worth investigating relations of contrast expressed through logical metaphor, in order to determine whether the comparatively low frequency of French interclausal and intersentential CMs of contrast is compensated for at lower levels of constituency.

     In summary, the results presented in this section have provided strong empirical evidence against the claims which have so far prevailed in the English-French contrastive literature. One caveat must be added with respect to these findings, however. Table 9 reports the frequencies of *individual* conjunctive markers of contrast identified in the corpus, taking no account of the fact that in some case, a single relation of contrast may have been signalled by two markers simultaneously, as in (11) and (12).

(11)   True, none are major cities. **But** the impact is **still** significant.
                                        (Mult-Ed-EN – The Independent)

(12)   C'est un dialogue fondé sur une méfiance mutuelle, **mais** c'est un dialogue
         quand même.   (Mult-Ed-FR – Le Figaro)

One undesirable consequence of this is that, if one of the two languages displays a greater propensity than the other to cumulate several markers to express a single relation of contrast, its degree of cohesive explicitness as compared to the other language will have appeared somewhat inflated by the counting method used here. In such a case, the language would not (simply) display a greater tendency to signal relations of contrast explicitly in discourse, but also/rather to be more emphatic when actually expressing these relations. That being said, French has been demonstrated to cumulate discourse-relational markers more frequently than English (see e.g. Crible 2018: 122). In addition, a pilot study carried out on a random sample of 300 English and 300 French CM tokens from the data set revealed no significant difference in frequency of combinations of CMs to express a single relation of contrast (i.e. 14/300 CM combinations of CMs of contrast in English, against 11/300 in French; χ^2=0.38; df=1; p=0.54). Therefore, it is likely that taking such a phenomenon into account would not have influenced the direction of the difference uncovered here.

7.2.2   Preferred types of conjunctive markers in English and French

This section presents the breakdown of the results across the three categories of conjunctive markers, viz. coordinators, subordinators and conjunctive adjuncts. The objective is to test the hypothesis that English will show a preference for coordinators, whereas French will tend to use subordinators and conjunctive adjuncts.

Importantly, while assertions to that effect are fairly widespread in the contrastive literature, it is not always entirely clear whether these preferences pertain to frequencies of use (e.g. more subordinators in French than in English), or proportions of use (e.g. larger proportion of coordinators out of the total number of CMs of contrast in English than in French). In this section, both the frequencies and proportions of use of each CM category are therefore discussed.

Table 10 provides the relative frequency of each category of contrastive CMs per language. The standardised residuals emerging from each chi-square comparison are provided between brackets next to the relative frequencies. The last column of the table provides the ratio of the frequency difference between English and French. Hash signs are used to indicate that the figure corresponds, either partly or entirely, to an extrapolation based on a random sample.[2]

Table 10. Breakdown of the relative frequencies (pmw) of English and French CMs per category

Types of CMs	English	French	Ratio of the difference
Conjunctive adjuncts	2,320 (+16.6)	1,342 (−16.2)	1.73
Coordinators	4,882# (+18.8)	3,233# (−18.4)	1.51
Subordinators	1,454 (+8.2)	1,054# (−8.0)	1.38
Total	8,656	5,629	1.54

What stands out from Table 10 is that the predominance of English over French in terms of CM frequencies uncovered in the previous section (see Table 9) holds across the board: for all three categories of markers, English displays a markedly higher frequency of conjunctive markers of contrast than French. The extent of this difference varies depending on the category, with the widest frequency gap being observed for conjunctive adjuncts, followed by coordinators, then subordinators. The difference is nonetheless statistically significant for all three classes of markers (CAs: $\chi^2 = 540.5$, $df = 1$, $p < 0.001$, $\varphi = 0.01$; coordinators: $\chi^2 = 693.8$,

---

2.   As a reminder, due to their very high frequency in the corpus, the English coordinator *but*, the French coordinator *mais*, and the French subordinator *si* were analysed using random samples of 2,000 occurrences (before disambiguation; see Chapter 5 for more details). To give a concrete example, out of the 11,876 instances of *but* retrieved in the Mult-Ed corpus, only a random sample of 2,000 instances were disambiguated manually. After disambiguation, only 1,650 of these 2,000 hits were found to meet the definition of conjunctive markers of contrast established in this book. This number was extrapolated in the following way: (1,650/2,000) * 11,876 [i.e. (number of occurrences kept/number of occurrences in the sample) * total number of hits in the corpus] = 9,798 relevant instances of *but* (raw frequency).

$df = 1$, $p < 0.001$, $\varphi = 0.01$; subordinators: $\chi^2 = 131$, $df = 1$, $p < 0.001$, $\varphi = 0.01$, with significant standardised residuals for each cell of the table). As a result, in terms of frequencies, the hypotheses formulated on the basis of the intuition-based literature are once again rejected, except for coordinators (viz. *mais* and *but*), which are indeed more frequent in English than in French. The results obtained for coordinators of contrast are also in line with those obtained by Cosme (2008a) who, looking at all formal and functional types of coordination in English, French and Dutch, uncovered a markedly lower frequency of coordinating structures in French newspaper editorials than in comparable English texts. For subordinators, the corpus results stand in contradiction with Cosme's findings, which revealed no significant difference in overall frequency of subordinating devices in English and French. It must nevertheless be underlined that Cosme's study was based on samples of 300 sentences per language, displaying a total of about 450 relations of subordination in English and French. This is a relatively small sample: as a matter of fact, the contrastive results obtained by Cosme for the French-Dutch language pair were challenged in a similar study carried out by Degand and Hadermann (2009). Whereas Cosme had found a significantly higher frequency of subordinate clauses in French than in Dutch, using the same measures of analysis in a comparable corpus of narrative texts, Degand and Hadermann found no significant differences in degree of syntactic complexity between the two languages. The findings presented in this chapter have a substantially greater power of generalisation than Cosme's, as they result from the analysis of nearly 5,000 subordinators of contrast (viz. c. 2,900 in English, and c. 1,800 in French).

While Table 10 has allowed us to identify very significant differences in frequency of markers across all three grammatical categories – with English systematically using more explicit markers of contrast than French – it is also worth wondering whether English and French differ regarding the types of grammatical resources which they tend to prefer when they actually signal contrast explicitly by means of a CM. The proportion of use of each type of CM out of the total number of conjunctive markers in the corpus is represented graphically in Figure 13. The bar chart also includes confidence intervals for each of the samples plotted on the graph.[3]

---

3. By definition, corpus analyses draw conclusions on language use on the basis of samples of data which are felt to be representative of a given language or language variety as a whole. While samples are supposed to provide a reliable estimate of what happens in the entire language (variety) investigated, it is also likely that the results obtained on the basis of a particular sample may not be exactly identical to those which would have been obtained if the entirety of that language (variety), or even if another, comparable sample had been analysed. Confidence intervals serve to delimit the range of values, around the results obtained for the sample(s)
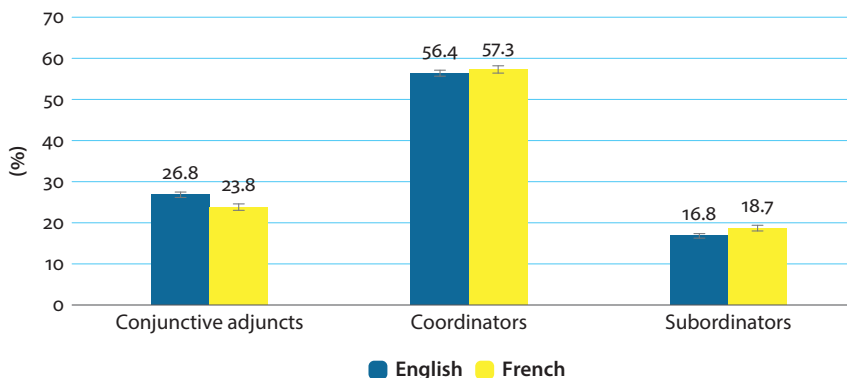
Figure 13. Proportion of use of each type of conjunctive marker in English and French (in percent)

As appears clearly from this figure, despite marked differences in frequencies of use of each type of marker, English and French display a certain degree of similarity regarding their preferred types of markers of contrast. In both languages, coordinators take the lion's share (viz. they represent more than half of the total number of markers), followed by conjunctive adjuncts (about a quarter of the occurrences). In English as well as French, subordinators are the least frequent type of marker to express a contrast between two clauses or sentences: they account for less than 20% of the total number of CMs identified in the corpus. Such results corroborate previous findings on CM usage. In his corpus-based study of contrastive linking in spoken and written English, for instance, Altenberg (1986) reports that coordinators make up 54% of all the contrastive markers identified in writing, against 25% for conjunctive adjuncts, and 21% for subordinators. Focusing on hypotactic and paratactic relational markers, Smith and Frawley (1983) and Nesbitt and Plum (1988), too, both report a significantly higher frequency of coordinating over subordinating devices across a range of English text types (fiction, news, religion and science in Smith and Frawley; narrative interviews in Nesbitt and Plum). Smith and Frawley (1983: 350) interpret these results as a manifestation of the 'naturalness' of coordination over subordination in language use, no matter how sophisticated the register. Interestingly, however, the results presented in Figure 13 stand in contradiction with the correlation established by Caffarel (2006: 54) between the number

---

analysed, within which the 'true value' is likely to fall (where 'true value' refers to the value that would have been obtained if the whole language (variety) had been analysed; see Field et al. 2012). In this case, the confidence intervals are fairly small, which suggests that the data analysed likely shows little variation with respect to English and French editorials as wholes.

of markers available in a given system, and the frequency of use of that system. For example, she notes that in French, relations of extension are more frequently realised through parataxis than hypotaxis, and attributes these results to the fact that there is a greater number of paratactic markers of extension than equivalent hypotactic markers. In the case of contrastive relations, however, the most frequent type of realisation (i.e. coordination) is also the one which offers the fewest lexical resources for expressing contrast, viz. a single CM type in each language (*but* and *mais*), against about 30 conjunctive adjuncts and a dozen subordinators of contrast per language. In this case, the high frequency of use of coordinators is probably rather due to the high degree of polysemy of these markers, as both *but* and *mais* have been shown to be the most polyfunctional markers of contrast within their respective language system (see e.g. Csüry 2001; Blakemore 2004).

While the ordering of the three categories is similar in the two languages (with coordinators being markedly more frequent than conjunctive adjuncts, and subordinators emerging as the least common type of realisation), a chi-square test of independence nevertheless reveals a significant difference between the English and French proportions of use of the three grammatical types of conjunctive markers ($\chi^2 = 40.5$, df = 2, p < 0.001, $\varphi = 0.04$). The standardised residuals, which indicate which of the categories give rise to significant differences between the two languages, are represented visually in an association plot in Figure 14.[4] Standardised residuals which indicate a non-significant frequency difference appear in grey on the plot. Residuals which point to a statistically significant difference (and are thus larger than the 1.96 significance threshold) appear in blue (positive residuals) and red (negative residuals; see Chapter 5 for more on how to interpret standardised residuals). The higher the boxes, the more the observed frequencies deviate from the expected frequencies. The width of the boxes reflects the prominence of the category in the total data set (e.g. in this case, the boxes for coordinators are the widest, as coordinators represent the most frequent category in both languages).

What Figure 14 shows is that, when it comes to explicitly signalling a relation of contrast between two discourse units, conjunctive adjuncts are chosen significantly more frequently in English than in French (as attested by the positive standardised residual in English and the negative residual in French, with both values topping the 1.96 significance threshold), while subordinators constitute a significantly more common choice in French than in English (cf. the negative standardised residual in English and the positive residual in French, again both

---

4.   All the association plots in this chapter are computed by means of the 'assoc' function from the 'vcd' R package (Meyer et al. 2017).
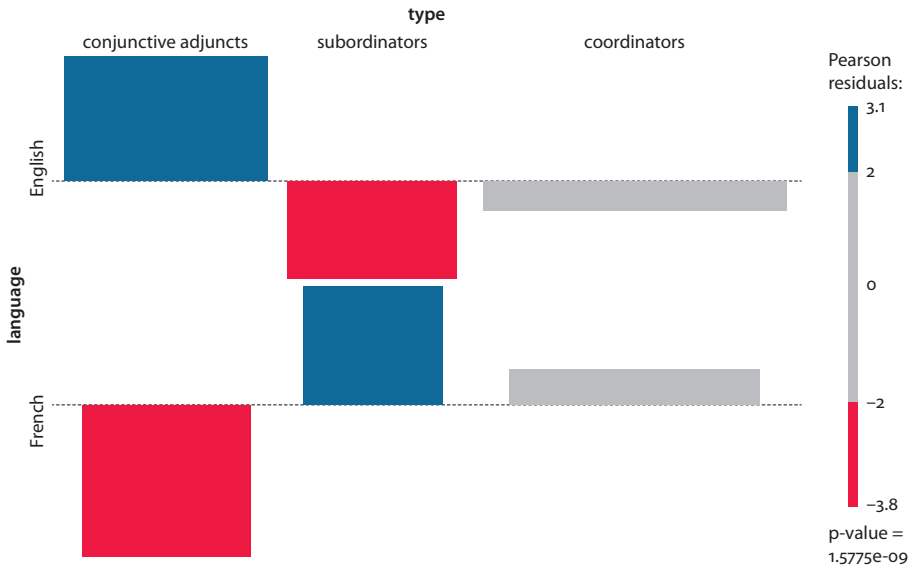
Figure 14. Standardised residuals for the proportions of use of English and French types of CMs

greater than 1.96.[5] The difference in proportion of use of coordinators, by contrast, is not significant (with both residual values being smaller than 1.96). The results obtained for both conjunctive adjuncts and coordinators thus, once again, stand in stark contrast with the hypotheses formulated on the basis of the literature, which predicted that conjunctive adjuncts would be more frequent in French, whereas coordinators would be more typical of English. The fact that subordinators are a more common choice in French than in English, on the other hand, is consistent with the initial hypothesis. This finding also partly corroborates Cosme (2008b: 344), who observed that one of the most frequent cases of translation correspondence between French hypotaxis and English parataxis corresponds to adversative clauses, as in (13) – where the use of an emphatic pronoun in addition to the CM in the French example also deserves notice.

(13)   EN. If Nato killed the innocent, it did so by accident; if Serbia killed the innocent, it did so on purpose.

FR. Si l'OTAN tuait des innocents, c'était par accident, **alors que** les Serbes, eux, les tuaient volontairement.   (Cosme 2008b: 344)

---

5.   The significance of these differences is also visible from the fact that in Figure 6.1, the confidence intervals do not overlap for these two categories.

A comparison of the results presented in Table 10 with those emerging from Figure 13 makes clear that frequencies and proportions of use are two distinct measures of comparison: the two yield partly different results concerning the preferred strategies used in English and French to express discourse relations of contrast. With respect to frequencies, English appears to use significantly more explicit CMs of contrast than French, whatever the grammatical category looked at. On the other hand, the analysis of proportions reveals that when it comes to explicitly signalling a relation of contrast by means of a CM, English chooses conjunctive adjuncts more frequently than French, which in turn uses subordinators more commonly than English. Coordinators are selected in a similar portion of the cases in the two languages. Table 11 summarises the results obtained through the application of each of these two measures, while also comparing them with the three hypotheses formulated at the beginning of the chapter.

Table 11. Preferred types of CMs of contrast in English and French: A summary

	CONJUNCTIVE ADJUNCTS	COORDINATORS	SUBORDINATORS
HYPOTHESIS	French > English	English > French	French > English
FREQUENCIES	English > French	English > French	English > French
PROPORTIONS	English > French	English = French	French > English
# OF MEASURES CONFIRMING THE HYPOTHESIS	0/2	1/2	1/2

The last line of the table indicates that, out of the three hypotheses considered here, one is simply rejected, i.e. conjunctive adjuncts of contrast are *not* more common in French than in English, either in terms of frequencies or proportions of use. The other two hypotheses, on the other hand, are confirmed by only one out of the two measures of comparison (viz. either frequencies or proportions). In addition, the table shows that five out of the six comparisons carried out in this section produced statistically significant differences. This provides support in favour of the statement that, despite having a similar set of cohesive resources at their disposal, different languages tend to make a distinctive use of these resources (see e.g. James 1980: 113; Halverson 2004: 562). That being said, the preference of both English and French for coordinators over conjunctive adjuncts, and for conjunctive adjuncts over subordinators, also sheds light on a certain amount of similarity between the two languages.

7.2.3   Lexical breakdown of the corpus results

Following the description of the overall frequencies of markers of contrast in English and French, and the analysis of the distribution of these markers across

the three grammatical categories of coordinators, subordinators and conjunctive adjuncts, this section adds one further degree of granularity to the discussion by providing the breakdown of the results per individual conjunctive marker in each subcorpus. It first provides a general comment on the lexical variety of CMs used in English and French, then zooms in on a few (pairs) of markers which are discussed in greater detail.

7.2.3.1   *Lexical variety of conjunctive markers in English and French*

Table 12 provides the breakdown of the corpus results per lexical item. The conjunctive markers are listed in decreasing order of frequency in the corpus. Figures that result from a random sample are followed by a hash sign. Interestingly, out of the 42 English and 50 French CMs included in the list compiled for the purposes of the corpus analysis, six English and nine French markers did not appear at all in the corpus (viz. English *anyways, by way of comparison, by way of contrast, oppositely, quite the contrary* and *quite the opposite;* French *à la place, au lieu, comparativement, en dépit que, malgré que, nonobstant, nonobstant que, quand bien même que* and *quand bien même* used as a conjunctive adjunct). In addition, nine English and fourteen French markers occurred less frequently than five times per million words, and are therefore not included in this discussion (viz. English *anyhow, by comparison, contrariwise, for all that, just the same, quite the opposite, regardless, then again,* and *whilst*; French *à l'opposé, a contrario, au lieu que, cependant que, d'un autre côté, d'autre part, de toutes (les) façons, de toutes (les) manières, en comparaison, par comparaison, par contre, à titre de comparaison, en même temps que* and *inversement*). In Table 12, I adopt a separatist approach, distinguishing between markers that are generally closely associated with each other (e.g. *though* and *although*; *on the other hand* and *on the other; au contraire* and *bien au contraire*). It was felt preferable to clearly differentiate between these variants, so as not to miss possible differences between them in terms of frequency of use or syntactic patterning, for example. *Nonetheless* and *none the less* are the only forms which are grouped, as they are considered as mere orthographic variants of one and the same marker. Note that, interestingly, there is exactly the same number (i.e. twenty-seven) of CMs topping the threshold of five occurrences per million words in English and French.

A first observation that can be made on this table is that, in the two languages, there is a very uneven distribution of the tokens across the twenty-seven marker types, with some CM types displaying a very high frequency in the corpus (e.g. *but, yet, however*; *mais, pourtant, si*), and other CMs being fairly uncommon in newspaper editorials (e.g. *whereas, in contrast; bien que, quand même*) – or even absent, as mentioned above. In other words, in both languages, the bulk of the cohesive work is performed by a small portion of the CM types. These results are

Table 12. Relative frequencies of English and French individual conjunctive markers of contrast (per million words)

	ENGLISH		FRENCH	
	CONJUNCTIVE MARKER	**RELATIVE FREQUENCY (PMW)**	**CONJUNCTIVE MARKER**	**RELATIVE FREQUENCY (PMW)**
1	But	4,882[#]	Mais	3,233[#]
2	Yet	771	Pourtant	491
3	However	602	Si/s'	313[#]
4	While	575	Même s(i)	296
5	Although	294	Alors qu(e)	261
6	Though (subo.)	252	Or	170
7	Even if	199	Cependant	135
8	Instead	184	Au contraire	120
9	Though (adv.)	163	En revanche	110
10	Still	152	Toutefois	65
11	Even though	86	Plutôt	62
12	Nevertheless	84	Tout en	55
13	Nonetheless/ none the less	66	Néanmoins	53
14	By contrast	45	Tandis que	51
15	Anyway	42	Tout de même	42
16	Even so	41	Bien qu(e)	30
17	On the other hand	39	Alors même qu(e)	26
18	Rather	29	De l'autre	22
19	Whereas	27	Malgré tout	21
20	Meanwhile	24	Quoiqu(e)	13
21	On the contrary	21	De toute façon	13
22	Albeit	21	Bien au contraire	11
23	At the same time	19	À l'inverse	10
24	All the same	11	Quand bien même	9
25	On the other	10	Quand même	8
26	In contrast	8	De toute manière	5
27	Conversely	7	En même temps	5

consistent with previous studies on CMs: Altenberg (1984; 1986), for instance, also reports great variation in frequency of the different CM types retrieved in two corpora of spoken and written English. Likewise, Cuenca and Bach (2007)

show that English, Spanish and Catalan alike tend to resort to one or two very frequent markers to express reformulation, whereas other reformulative expressions are used much more sparingly (i.e. less than ten occurrences in their corpus). In addition, the overwhelming importance of the coordinators *but* and *mais* in the corpus (more than half of the total number of tokens in each language) is also in line with previous research on CMs of contrast. In her study of contrastive markers in English, German, Spanish and Portuguese, Rudolph (1996: 315) stresses that in all four languages, about two thirds of the total number of occurrences of contrastive CMs are taken up by *but, aber, pero* and *mas*, respectively. Altenberg (1986: 15) also notes that English *but* represents over 50% of the CM tokens retrieved in a corpus of informative prose. With respect to the French language, Csüry (2001: 120) reports that *mais* represents no less than 70% of the total number of CMs of contrast identified in a written corpus made up of literary, journalistic and institutional written texts, with the other 21 CMs of contrast sharing the remaining 30% of the occurrences.

In order to get a more precise picture of the variety of CMs used by the two languages to express contrast, the cumulative percentages represented by the various markers in the corpus are provided in Table 13. Each line of the table indicates what proportion of the total number of CM tokens is represented by a given (group of) CM type(s).

Table 13. Cumulative percentages of English and French CMs of contrast

	ENGLISH		FRENCH	
	CONJUNCTIVE MARKER	CUMULATIVE PERCENTAGE	CONJUNCTIVE MARKER	CUMULATIVE PERCENTAGE
1	But	56.4%	Mais	57.4%
2	Yet	65.3%	Pourtant	66.2%
3	However	72.3%	Si/s'	71.7%
4	While	78.9%	Même s(i)	77%
5	Although	82.3%	Alors qu(e)	81.6%
6	Though (subo.)	85.2%	Or	84.6%
7	Even if	87.5%	Cependant	87%
8	Instead	89.6%	Au contraire	89.2%
9	Though (adv.)	91.5%	En revanche	91.1%
10	Still	93.3%	Toutefois	92.3%
11	Even though	94.3%	Plutôt	93.4%
12	Nevertheless	95.2%	Tout en	94.3%

(*Continued*)

Table 13. (*Continued*)

	ENGLISH		FRENCH	
	CONJUNCTIVE MARKER	CUMULATIVE PERCENTAGE	CONJUNCTIVE MARKER	CUMULATIVE PERCENTAGE
13	Nonetheless/ none the less	96%	Néanmoins	95.3%
14	By contrast	96.5%	Tandis que	96.2%
15	Anyway	97%	Tout de même	96.9%
16	Even so	97.5%	Bien qu(e)	97.4%
17	On the other hand	97.9%	Alors même qu(e)	97.9%
18	Rather	98.3%	De l'autre	98.3%
19	Whereas	98.6%	Malgré tout	98.7%
20	Meanwhile	98.9%	Quoiqu(e)	98.9%
21	On the contrary	99.1%	De toute façon	99.1%
22	Albeit	99.3%	Bien au contraire	99.3%
23	At the same time	99.6%	À l'inverse	99.5%
24	All the same	99.7%	Quand bien même	99.7%
25	On the other	99.8%	Quand même	99.8%
26	In contrast	99.9%	De toute manière	99.9%
27	Conversely	100%	En même temps	100%

As Table 13 makes clear, the figures obtained in English and French are uncannily similar: in both languages, a single CM type (viz. *but* and *mais*) accounts for nearly 60% of the tokens, and two markers (i.e. *but* + *yet* and *mais* + *pourtant*) make up about 65% of the total number of occurrences in the two subcorpora. Interestingly, the first two pairs of English and French markers (i.e. *but/mais* and *yet/pourtant*) are made up of items which are generally viewed as translation equivalents (see below for a more detailed discussion). Moreover, in English as well as French, the 80% threshold is reached with the fifth most frequent marker (viz. *although* and *alors que*), and nine markers represent about 90% of the total number of tokens. In other words, in both languages, one third of the CM types (i.e. 9 CMs) represent 90% of the CM tokens. The great similarity of the English and French lexical distributions is represented graphically in Figure 15, where the English cumulative percentages are represented by bars, whereas the French data is represented by a curve. As appears clearly from the graph, the two distributions overlap nearly perfectly, so that each bar ends almost exactly where the curve passes.

More generally, what both Table 13 and Figure 15 show is that the size of the frequency difference between English and French is (more or less) constant throughout all the ranks of the list. The difference in overall frequencies between
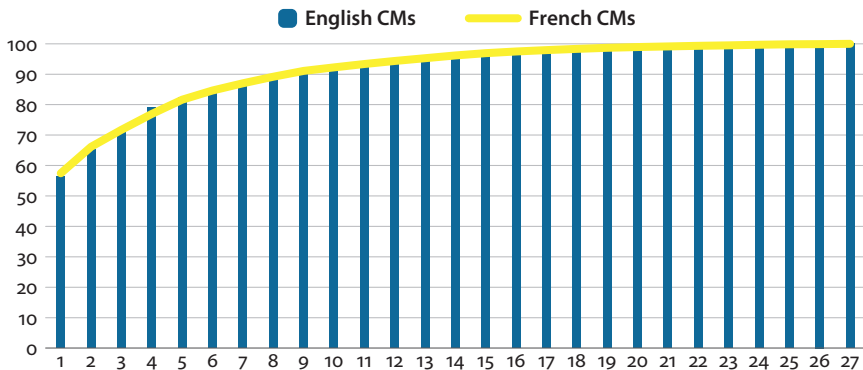
Figure 15. Cumulative percentages of the English and French conjunctive markers of contrast in the Mult-Ed corpus

English and French does not appear to be due to the very frequent use of one or two markers in English, but rather seems to stem from a generally more frequent use of most English CMs of contrast as compared to their French counterparts. As a result, the ranks of the English and French CMs (e.g. second, third, seventh, etc. most frequent CM) and the cumulative proportions represented by these ranks are roughly identical in English and French. The results reported in Table 13 are some-what surprising, as they stand in direct opposition with previous contrastive results described in the literature. Cuenca and Bach (2007: 153), for instance, explain that Spanish and Catalan display nearly twice as much lexical variety than English in the expression of reformulation (see also Cuenca 2003 for similar results). Like-wise, Kunz and Lapshinova-Koltunski (2014: 251) note a markedly higher vari-ety of conjunctive markers in German than in English written language. Finally, Wang (2011: 56) observes that, although about twenty CMs of contrast account for around 95% of the total number of contrastive CMs in English and Chinese alike, the distribution of these twenty markers is much more uneven in English than in Chinese. In English, a single marker (i.e. *but*) covers more than half of the tokens, against only about 20% for the most frequent Chinese marker, i.e. *dàn*.

It is worth noting that the results presented in Tables 12 and 13 may be of great interest for foreign language teaching. One of the approaches put forward by Crewe (1990: 321) to improve learners' usage of conjunctive markers is to encour-age them to focus on small sets of markers and attempt to achieve full mastery of their distinctive properties and usage values. This method contrasts with usual teaching strategies, which consist in presenting learners with long, undifferenti-ated lists of CMs without taking the time to present them with the specific features of these markers (ibid.: 323). The information presented in Tables 12 and 13 may emerge as a valuable tool to select the CMs which are most worthy of pedagogical

attention, by virtue of being the most widespread items in the learners' target language. Thus, for example, it would make more sense for English language teachers to teach their students how to use markers such as *but, yet, however* or *while* appropriately, without devoting too much time (if any) to less frequent markers such as *all the same* or *conversely*, which would make a much smaller contribution to improving the learners' writing. In fact, if anything, drawing the learners' attention to the CMs occurring at the bottom of the list may make those markers much more prone to overuse by learners, who are usually not fully aware of the preferred forms used by native speakers to express a given discourse relation. Naturally, the ranking of the CMs may also vary in function of the register investigated, a fact to which learners should also be sensitised. This issue will be tackled in Section 7.4.

7.2.3.2   *Frequency features of individual conjunctive markers of contrast*

Alongside general observations on the lexical variety of contrastive markers in the two languages, Tables 12 and 13 also allow for more focused comments on CM use in English and French. As pointed out on several occasions, it is very difficult to establish clear mappings between CMs across languages, even when they express the same type of meaning (see e.g. Degand 2004; Zufferey & Cartoni 2012; Hasselgård 2014a). This is especially true as the present study does not offer any fine-grained semantic analyses of the markers investigated, which severely restricts the type of correspondences that can be made between them. For these reasons, the discussion will be limited to a few pairs of markers which are very commonly presented and/or perceived as cross-linguistic equivalents – including in bilingual dictionaries. It is divided into four main categories: (i) perceived equivalents which occupy a similar position in the English and French systems; (ii) perceived equivalents which display markedly different frequencies in the two subcorpora; (iii) frequent CMs in one language which do not display any direct equivalent in the other language; (iv) stylistic comments.

A first observation emerging from Tables 12 and 13 is the striking similarity between English and French in the top cells of the tables: the first and second ranks are occupied by the two pairs of perceived equivalents *mais*/*but,* and *yet*/*pourtant*, with the members of these pairs accounting for roughly the same proportion of the CM tokens in the corpus. English *but* and *yet* on the one hand, and French *mais* and *pourtant* on the one other, together account for about 65% of the tokens in each subcorpus, which means that the bulk of explicit contrastive relations tend to be signalled by roughly equivalent devices in English and French editorials. Such clear frequency correspondences between items which are generally perceived as semantico-functional equivalents are not found anywhere else in the tables.

Conversely, the results also reveal that some markers which are generally perceived as equivalent hold very different ranks in the English and French systems

of contrastive linking. In this respect, some results reflect tendencies which have already been noted in the literature. This is the case of the cognate pair *on the contrary/au contraire*, whose members occur with markedly different frequencies in the corpus. These results are fully in line with Lewis (2006), who also reported a significantly lower frequency of English *on the contrary* than French *au contraire* in a comparable corpus of political speeches. For Lewis (ibid.: 12), this frequency difference reflects the partial nature of the semantic equivalence between the two markers, with the meaning of *on the contrary* being more restricted than the meaning of *au contraire*. She notes that, whereas *au contraire* "is often found in contexts of even-handed contrast between two ideas, […] *on the contrary* seems restricted to the more argumentational contexts in which the speaker/writer rejects the first idea as false and replaces it with the preferred, second, true idea" (ibid.). This situation is also observed in the Mult-Ed corpus. All the occurrences of *on the contrary* in the corpus correspond to the contrastive subtype of correction, where the second segment in the sequence serves to correct a piece of information that was negated in S1 (see Example (14)). By contrast, French *au contraire* can be used to express both correction – as in (15) – and opposition, as in (16), where the marker simply indicates a difference between the two related segments (viz. in terms of the consequences which the Debré and the Stasi projects might have). Note that this semantic difference between *on the contrary* and *au contraire* most probably explains the striking overuse of *on the contrary* by French-speaking learners of English (viz. 38 occurrences per 100,000 words, against 4 occurrences in native English), as reported by Granger and Tyson (1996: 21).

(14)   It is not the case that pensioners get a good deal in modern Britain. **On the contrary**, we still treat the elderly in a miserly, begrudging fashion.
(Mult-Ed-EN – The Independent)

(15)   En un an de pouvoir, Mohamed Morsi n'a pas su faire la preuve qu'un islam modéré peut évidemment s'inscrire dans un cadre démocratique. Il a, **au contraire**, suivi sa propre dérive autoritaire, focalisant toutes les frustrations autour de sa réforme controversée de la Constitution, qui porte clairement atteinte aux libertés publiques et aux libertés religieuses.
(Mult-Ed-FR – Libération)

(16)   Le récent projet Debré, d'interdire tout signe à l'école, même privée, constituait, lui, une négation du fait religieux. Le projet Stasi risque **au contraire** de l'exacerber.   (Mult-Ed-FR – Le Figaro)

The results also draw attention to other striking frequency gaps between CMs which are often presented as cross-linguistic equivalents. One such case concerns the pair *however/cependant*. Although these markers are nearly always presented as primary translation equivalents in bilingual dictionaries, in the Mult-Ed

corpus, *however* is markedly more common than its perceived French equivalent *cependant*. While *however* is the third most frequent CM in the English subcorpus, *cependant* only occupies the seventh rank within the set of French CMs of contrast. As was the case for *on the contrary/au contraire,* one possible explanation for this phenomenon may pertain to a difference in degree of polysemy of the two markers. As highlighted by Prasad et al. (2008: 67), although English *however* predominantly expresses concession, it may also be used with a meaning of opposition. A cursory glance at the corpus data confirms that *however* is indeed sometimes used to express opposition in the Mult-Ed corpus, as in (17) and (18), where the CM could easily be replaced by *by contrast*. In (18), for instance, *however* is used to simply state a difference between the fashions of central banking (i) in the 2000s and (ii) after the financial crisis. On the other hand, the semantic coding of a fifty-occurrence random sample of French *cependant* in Mult-Ed did not shed light on any clearly oppositive uses of the marker. Instead, *cependant* was found to be mainly used in concessive situations – as in Example (19), where its meaning rather corresponds to that of *yet* (see also Dupont 2013: 139–154 for a comparison of the semantic properties of *cependant* and *however*). It is clear, however, that a more fine-grained analysis of the semantic properties of these two markers would be required in order to establish any definite conclusions on their diverging frequencies in the corpus.

(17)   [U]nless it can be shown that Mr Foster used the Prime Minister's name to reduce the purchase price, which remains unproven at this stage, this investment is a curious but not a reprehensible one. What was unacceptable, **however**, was the ruthless and dishonest operation – on the Prime Minister's behalf – to try to squash the story.

(Mult-Ed-EN – The Independent)

(18)   Just like many other areas of policy making, central banking has its fashions too. Throughout the 2000s, the world's largest monetary authorities preferred to retain as much flexibility as possible over the future path of interest rates […] After the financial crisis, **however**, it became more common for the monetary authorities to commit to a pre-established strategy.   (Mult-Ed-EN – The Financial Times)

(19)   Les Français juifs sont nombreux, et certains sont enracinés chez nous depuis des lustres. Ils sont in-té-grés, et, **cependant**, ils sont de plus en plus nombreux à estimer que la France devient antisémite, et j'en connais qui font vivre leurs enfants, petits ou grands, dans un pays dont on ne doit pas exclure qu'on pourrait le quitter.   (Mult-Ed-FR – Le Monde)

Finally, the very high frequency of the French subordinator *si* in the corpus also points to another case of partial equivalence between English and French CMs.

Si is the third most frequent marker (and the most frequent subordinator) in the French data, with a frequency of 313 per million words. In Mult-Ed, *si* assumes meanings of both opposition – as in (20), where *si* could be translated by *whereas* – and concession – as in (21), where *si* broadly corresponds to English *although*.

(20)   **Si**, en Italie, tout le monde ou presque trouve normal le renvoi chez lui de l'ancien terroriste, en France, beaucoup s'indignent.

<div align="right">(Mult-Ed-FR – Le Figaro)</div>

(21)   Or **si** elle reste, avec 1,2 milliard de pratiquants, la première religion de la planète, l'Eglise catholique, à bien des égards, est mal en point.

<div align="right">(Mult-Ed-FR – Le Monde)</div>

Therefore, the French marker *si* appears to play a key role in the expression of contrastive links in written language. This result is in line with Lewis (2005), who also stressed the relative importance of concessive *si* in political discourse (with a frequency of 254 occurrences per million words). Unlike French *si*, English *if* is not commonly used to express a meaning of contrast but tends to be restricted to conditional uses. In a random sample of 200 occurrences of *if* extracted from Mult-Ed, I only found three instances of English *if* which could be interpreted as contrastive (see also Lewis 2005: 40 for a similar observation). One such example is provided in (22). In other words, only a portion of the uses of French *si* – viz. mostly, its conditional uses – are covered by those of English *if.*

(22)   In just four days two of the six founding members of the then European Economic Community [viz. France and the Netherlands], forerunner of today's union of 25 countries and 455 million people, have rejected a document designed to meet the challenges of the continent's future. **If** the results were similar in France and the Netherlands, different factors were at work.   (Mult-Ed-EN – The Guardian)

This partial semantic equivalence between French *si* and English *if* might be worth some pedagogical attention, as it is likely to cause French learners of English to misuse the subordinator *if,* generalising the polysemy of French *si* to its perceived equivalent in English. This is especially true as French *si* seems to hold a central place in the system of contrastive linking of written French. In fact, evidence of negative transfer in French-speaking learners of English with respect to *if* is provided by Zufferey et al. (2015), who show that this learner population scored significantly lower than both native speakers and Dutch-speaking learners of English at an experimental task requiring them to spot incorrect (viz. contrastive) uses of *if* among a sample of fifty-five sentences. Consequently, it may be of interest to draw learners' attention to English subordinators which can actually be used as equivalents to the contrastive uses of French *si*. In this respect, *while* appears to be a good

candidate, since it can also convey both opposition – as in (23) – and concession – as in (24). In addition, *while* takes on a similar importance to French *si* in the English system of cohesive resources, being the fourth most frequent English CM of contrast (whereas *si* is the third most frequent CM in French).

> (23)   **While** the birds, butterflies and bees who enjoy our plants are a source of pride, slugs are fair game for fear, loathing and extermination.
>
> (Mult-Ed-EN – The Guardian)

> (24)   **While** it may be true that these are, as the White House and Number 10 insist, only isolated incidents, they again point to the familiar dangers of using armed forces as peacekeepers, a role for which they are rarely trained or equipped to perform.   (Mult-Ed-EN – The Guardian)

Conversely, it is also important to draw the attention of English (and other) learners of French to the potential of French *si* for contrastive linking. Contrastive uses of *si* are arguably not very salient in learners' minds, as the CM is most often associated with conditionality. A survey of five grammars of French as a foreign language aimed at intermediary or advanced learners revealed that *si* is never mentioned as one of the devices that can be used to express opposition and concession.[6] Instead, some of the markers which are almost systematically mentioned include *bien que, quoique, alors que, tandis que, pourtant, mais, cependant, quand même, à l'opposé, par contre,* etc. (see e.g. Grégoire & Thievenaz 2017: 244; Boularès & Frérot 2017: 152; Bourmayan et al. 2017: 125; Glaud et al. 2015: 178; Gliemann et al. 2015: 222). Yet, in the Mult-Ed corpus, most of these markers appear much more infrequently than *si* to signal a contrastive link: as appears from Table 12, for instance, *tandis que, bien que* and *quoique* only display a frequency of 51, 30 and 13 occurrences per million words, respectively, against 313 for *si* (even though these frequencies may of course differ in a corpus representing another language register).

Other particularly striking frequency differences between English and French CMs concern the two pairs *(al)though/bien que* and *instead/plutôt*. Again, the markers in these two pairs are very frequently presented as translation equivalents in English-French bilingual dictionaries.[7] Yet, as appears from Table 12, *plutôt* (62 pmw) is three times less frequent than its assumed equivalent *instead*

---

6.   Viz. (i) *Grammaire progressive du français: Niveau intermédiaire*; (ii) *Grammaire progressive du français: Niveau avancé*; (iii) *Grammaire essentielle du français: Niveau B2*; (iv) *Grammaire essentielle du français: Niveau B1*; and (v) *Focus: Grammaire du français*.

7.   See for example *Le Robert & Collins anglais-français*, or *Larousse anglais-français en ligne* (<http://www.larousse.fr/dictionnaires/francais-anglais> (12 April 2020)).

(184 pmw). Likewise, *bien que* (30 pmw) is markedly less frequent than its two putative counterparts *though* (252 pmw) and *although* (294 pmw). In the Mult-Ed corpus, the role of *although* and *though* seems to be assumed by French *alors que.* All in all, the frequency discrepancies emerging from the corpus can therefore be said to provide confirmatory evidence in favour of the following statement, made by Chuquet and Paillard (2017: 38):

> Despite the apparent resemblance of the [English and French] reservoirs of con-nectives, these elements are not necessarily used with the same frequency, or in the same contexts. Therefore, it can be misleading to establish a systematic corre-spondence between two connectives (such as those that are sometimes suggested by bilingual dictionaries) [my translation].

Alongside (i) perceived equivalents that hold a similar position in the French and English systems (e.g. *mais/but*), and (ii) perceived equivalents which appear to have unequal importance in the two languages (e.g. *however/cependant*), a third category of corpus results pertains to markers which do not seem to have any clear or direct equivalent(s) in the other language, but occur with a rather high frequency in the corpus. The most striking example of this phenomenon is the French marker *or*, whose exact meaning is notoriously difficult to render in other languages (see e.g. Grieve 1996: 363; Rey 1999; Chuquet & Paillard 2017: 41). French *or* has no direct translation equivalent in English and, accordingly, is often removed in translation. Chuquet and Paillard (2014: 41), for example, report that out of 32 occurrences of sentence-initial *or* in a corpus of French newspaper arti-cles, two thirds (i.e. 21) simply give rise to a juxtaposition of sentences in English, as in (25). Other translations uncovered by Chuquet and Paillard (ibid.) include *but* (7/32) and *and* (3/32).

(25) FR. Pour que la proposition française soit retenue, il faudrait donc qu'elle soit acceptée par les 28 autres négociateurs. **Or**, aux dernières informations, elle aurait, au mieux, l'appui de 5 ou 6 délégations.

EN. In order to be adopted, the French proposal would have to be accepted by the other 28 negotiators. At the last count it could rely on the support of five or six delegations at most.   (Chuquet & Paillard 2017: 41)

French *or* is rather frequent in the French subpart of Mult-Ed: it is the sixth most widespread marker in the corpus, with a frequency of 170 per million words. Interestingly, the observation that such a frequent French marker has no direct equivalent in English – thus leading to a high ratio of zero translations in the French-English translation direction – may have partly contributed to the wide-spread claim that French tends to be more explicitly cohesive than English overall.

One meaning of *or* that appears frequently in the French subcorpus combines a concessive meaning with a syllogistic one. A syllogism is a kind of deductive

reasoning in which the conjunction of two accepted facts – called premises – is presented as leading to an incontestable conclusion, as in the oft-cited Example (26). In French, the marker *or* would occur between the two premises, i.e. between the first and the second clauses. The conclusion itself may but must not be expressed explicitly. Examples of *or* expressing a mixed concessive and syllogistic relation are provided in (27) and (28).

(26)   All men are mortal. Socrates is a man. Therefore, he is mortal.

(27)   Deuxième argument : la technique conduit à détruire des cellules embryonnaires à différents stades de développement, **or** tout embryon n'est-il pas une «personne potentielle» ?   (Mult-Ed-FR – Le Figaro)

(28)   La raison principale de l'arrivée massive d'étrangers à l'entrée du tunnel est, on le sait, l'accueil plus favorable que leur réserve la Grande-Bretagne. **Or** celle-ci, comme elle s'y était engagée, vient de durcir sa législation pour cesser d'apparaître comme une terre promise et décourager, autant que possible, les candidats à l'immigration. Quand cette nouvelle se sera répandue dans leurs pays d'origine, beaucoup renonceront sans doute à entreprendre le long voyage vers le nord de la France.

(Mult-Ed-FR – Le Monde)

Example (27) is part of a series of arguments in favour of banning therapeutic cloning. The relation between the two segments can be interpreted in two main, coexisting ways: on the one hand, opponents of therapeutic cloning are building a syllogistic argumentation, by stating that since therapeutic cloning consists in destroying embryonic cells *and* since these very cells can be considered to be potential people, cloning is immoral. On the other hand, the example could also be read in the following way: despite the fact that all embryos are potential people, some scientists are destroying them for the purpose of therapeutic cloning. A similar argumentative development is at play in (28), where (i) S1 and S2 are both given as arguments towards the conclusion that immigrants will stop coming to France as soon as they hear that it is no longer possible to cross the channel; and (ii) a contrast can also be perceived between the fact that immigrants come to France with a view to reaching the UK, and the fact that, because of the new legislation, crossing the tunnel will no longer be possible. Whereas this mixed concessive and syllogistic type of meaning is typical of French *or*, English does not seem to have any marker that is specifically tailored to express that meaning. This may explain the high frequency of omissions of *or* in translation from French into English.

Conversely, English editorials display a high frequency of the conjunctive adjunct *still* (152 pmw), which does not have any clear equivalent in French. As explained in Chapter 6, English *still* frequently expresses a combined contrastive

and temporal meaning, as in (29), where S2 could be paraphrased either as "younger voters turned up in sufficient numbers anyway", or as "the number of younger voters remained sufficient" (a reading which is further supported by the following verbless clause "No longer").

(29)   The young, especially the less well-educated, have always been more
        indifferent to voting than their parents and grandparents. In a world where
        turnouts in general elections were once as high as 80 per cent – as was
        the case when today's pensioners first cast a vote for Attlee or Churchill –
        that did not matter so much. Younger voters **still** turned up in sufficient
        numbers to make their voice heard. No longer.

                                                      (Mult-Ed-EN – The Independent)

In French, no frequent conjunctive adjunct allows for such a subtle combination of temporality and contrast: instead, when they link two sentences by means of a conjunctive adjunct, French writers need to choose between a marker of contrast (e.g. *quand même*) or a temporal signal (e.g. *encore, toujours*) – or, possibly, resort to a combination of two markers. Yet, according to Gallagher (1995: 205), when relations of both contrast and temporality hold simultaneously between two clauses or sentences, many French writers tend to favour the explicit marking of temporal relations over contrastive ones – another preference which may help account for the overall gap in frequency of contrastive CMs between English and French.

    Finally, from a stylistic perspective, the relatively common use of the English conjunctive adjuncts *though* (163 pmw), *still* (152 pmw) and *anyway* (42 pmw) in written texts – exemplified in (30) to (32) – is also startling. These three markers have usually been associated with spoken language or informal types of discourse (see e.g. Biber et al. 1999: 886; Quirk et al. 1985: 636; Rudolph 1996: 327). It is therefore somewhat surprising to find them in quality newspaper editorials, a written register which has been said to rank relatively high in terms of formality (Biber 1988: 128; Granger 2014: 62).

(30)   No doubt, all would find the reality of government tougher than the theory
        of opposition. What is clear, **though**, is that the challenge is the same
        whether these countries elect new leaders or stick with the old ones.

                                                      (Mult-Ed-EN – The Economist)

(31)   The process has seemed complicated, drawn-out and fraught with political
        horse-trading. **Still** it is to be welcomed.   (Mult-Ed-EN – The Guardian)

(32)   Nor is there much reassurance to be found in the fact that the AK Party's
        leader, Recep Tayyip Erdogan, a former Mayor of Istanbul, is not running
        himself after being jailed for reading an aggressive religious poem. He
        might end up as Prime Minister **anyway**, as he remains an obvious
        candidate for the job.   (Mult-Ed-EN – The Times)

Such results are probably attributable to an evolution of English newspaper prose in general, and newspaper editorials in particular, towards an increasingly conversational and colloquial style since the second half of the twentieth century (e.g. Mair & Hundt 1995; Hundt & Mair 1999; Westin 2002: 39, 151; Steen 2003). In her diachronic corpus-based analysis of English newspaper editorials, for example, Westin (2002: 39) uncovers a significant increase, over the twentieth century, of linguistic features generally associated with conversational discourse, including questions, *not*-negations, imperatives and contractions. She explains that this evolution is indicative of a "drift [of newspaper editorials] not only towards more informal language but also towards a more conversational type of informal language" (ibid.). In French, on the other hand, none of the CMs appears to be particularly informal.

7.3   Syntactic patterning of conjunctive markers of contrast

The objective of this section is to describe the syntactic features of the segments in which English and French conjunctive markers of contrast tend to be included. Such a syntactic approach to CMs is not common in research on discourse relations. While some studies on conjunctive markers have included reflections on syntax, the objective has typically been to draw the line between the units which may and may not be viewed as segments of discourse (see e.g. Hoek et al. 2017; Muller et al. 2012). In other words, the meeting point between syntax and the study of discourse relations has usually been restricted to methodological issues pertaining to the segmentation of texts into discourse units – such as those tackled in Chapter 6. The syntactic features of the segments related by conjunctive markers have rarely been investigated in their own right. In fact, I am only aware of a single study that has systematically identified the syntactic features of the discourse segments related by CMs, viz. Altenberg's (1984) corpus-based analysis of causal markers in English speech and writing. To my knowledge, no study of this type has been carried out in a cross-linguistic perspective. This section seeks to fill this gap in current contrastive research on CMs, by consistently annotating the syntactic features of the clauses hosting English and French CMs of contrast. The syntactic features of these clauses will be defined along two main dimensions, viz. clause types (i.e. finite, non-finite, verbless) and rank status (i.e. main, hypotactic, embedded or minor). For reasons of feasibility, the analysis focuses exclusively on the features of the host clause, corresponding to the second segment in the relation (or S2), which is introduced by the marker.

Such an approach to CMs may appear surprising at first sight. Purely syntactic categories such as finite or verbless clauses are typically not considered to be

significant for the study of cohesion. More generally, a fairly clear-cut distinction has usually been made between the syntactic/grammatical and the discourse levels of the linguistic system. This is epitomised by Halliday and Hasan's (1976) initial account of cohesion, where the authors draw a very strict dividing line between relations of grammatical dependence, on the one hand, and cohesive relations, which are completely ungoverned by structural considerations, on the other (see Chapter 2 for a discussion). This section seeks to demonstrate the relevance of integrating syntactic analyses into the study of conjunctive markers. As will be shown, for example, such an approach sheds light on some subtle strategies of textual development in English and French editorials, revealing that certain syntactic uses of CMs may create specific stylistic effects, such as emphasis or compression. The approach also uncovers a certain degree of interaction between syntax and lexis when it comes to CM use, by underlining distinctive syntactic features, within each language system, of CMs belonging to the same grammatical category (e.g. *while* does not display the same syntactic tendencies as *whereas*). Moreover, the fine-grained syntactic coding of the host segments also makes it possible to take account of the fact that, as highlighted by a number of researchers including Lehmann (1988), Matthiessen (2002) or Cosme (2008a), it is counter-productive to conceive of coordinators, subordinators and conjunctive adjuncts as either fully homogeneous or clearly distinct categories of markers. Rather, as will also be shown in this section, each category of markers encompasses a variety of possible uses (e.g. finite vs. non-finite subordination) representing more or less explicit and/or integrated ways of expressing a relation between two discourse units. This diversity of uses within each CM category results in a blurring of the boundaries between the three classes of markers, which are therefore best conceived of as different levels on a cline.

This section is made up of three main subparts. Section 7.3.1 provides a descriptive account of the syntactic features of English and French conjunctive markers of contrast. Section 7.3.2 focuses on some discourse effects associated with certain syntactic uses of CMs of contrast, thus demonstrating that CM use is partly situated at the syntax-discourse interface. Finally, Section 7.3.3 discusses the influence of lexis on the features of the host clauses.

7.3.1   Syntactic patterning of English and French conjunctive markers of contrast

This section presents the quantitative results emerging from the syntactic coding of the segments introduced by the English and French CMs of contrast in Mult-Ed. The three categories of CMs are discussed separately, as the type of marker investigated clearly influences the form taken by the host clause (e.g. a subordinator will

most likely introduce a hypotactic clause). Note that the discussion in this section will remain mainly descriptive. It is intended to set the scene for the following sections, which provide more interpretive and/or qualitative perspectives on these quantitative results.

7.3.1.1   *Syntactic patterns of English and French conjunctive adjuncts of contrast*
Table 14 presents the syntactic features of the clauses in which English and French conjunctive adjuncts of contrast are included. As appears clearly from these figures, in both English and French, an overwhelming majority of CAs of contrast are used in (i) finite and (ii) main clauses, as in Examples (33) and (34), where the finite verb forms are underlined.

(33)   At Nansen, there were 'no lessons in the humanities, arts or music' for an entire year group. Arabic, **however**, was compulsory learning across the entire school.   (Mult-Ed-EN – The Telegraph)

(34)   L'exercice du pouvoir, surtout quand celui-ci est octroyé démocratiquement, implique des responsabilités dont les contours doivent être clairement définis. **Pourtant**, les citoyens s'y perdent.   (Mult-Ed-FR – Libération)

Table 14. Syntactic patterning of English and French conjunctive adjuncts of contrast

Syntactic Features of the Host Clause		English	French
Clause Type	Finite	97.9%	90.7%
	Non-Finite	1.7%	4%
	Verbless	0.4%	3.3%
	Phrasal-Clausal	0%	2%
Rank Status	Main	95.1%	82.6%
	Hypotactic	2.5%	9%
	Embedded	1.8%	2.5%
	Minor	0.5%	3.8%
	Phrasal-Clausal	0%	2%
Total nb. of Occurrences (Raw)		4656 (100%)	2812 (100%)

The strong predominance of the finite and main patterns in the corpus is not surprising, since these are the default syntactic categories for both English and French clauses. In spite of this shared dominant pattern, however, the corpus results also point to a number of significant differences between English and French CA usage. In fact, chi-square tests of independence returned significant

differences between English and French for both clause types ($\chi^2 = 241.9$, $df = 3$, $p < 0.001$, $\varphi = 0.18$) and rank status ($\chi^2 = 398.2$, $df = 4$, $p < 0.001$, $\varphi = 0.23$). The standardised residuals for these two tests are provided in Figures 16 and 17, respectively. These allow us to go beyond the single p-values obtained for each contingency table and identify what levels of the 'clause type' and 'rank status' variables give rise to significant differences between English and French. As appears from these two figures, all but one syntactic category (viz. embedded clauses) contain at least one standardised residual topping the significance threshold of 1.96. This means that, apart from embedded clauses, all the syntactic categories included in the analysis give rise to significant frequency differences between English and French. More specifically, the figures show that the dominant finite and main patterns are more likely to occur in English than in French (where they display significantly negative residuals), whereas the opposite is true of all the other categories.
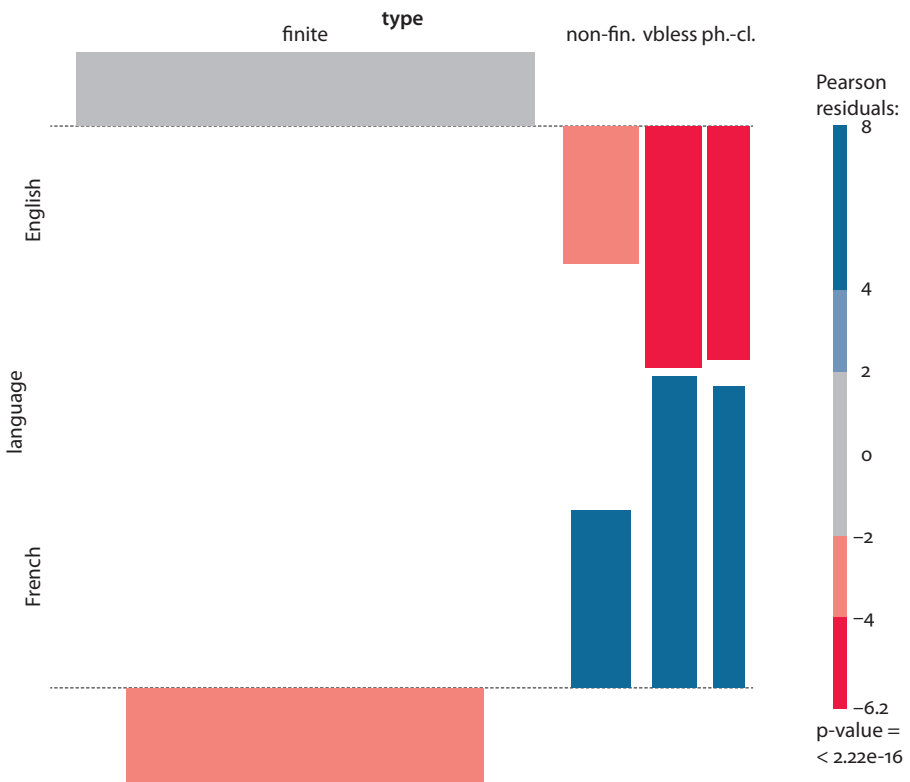
Figure 16. Standardised residuals for the types of clauses hosting English and French conjunctive adjuncts
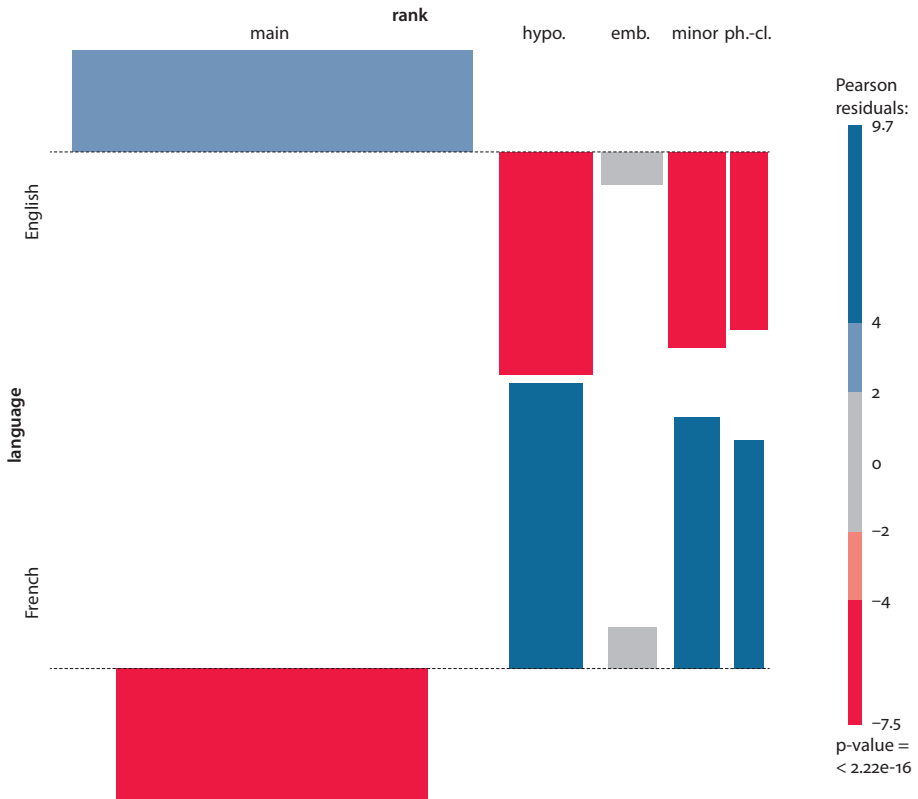
Figure 17. Standardised residuals for the rank status of clauses hosting English and French conjunctive adjuncts

Firstly, in terms of clause types, the results reveal that, although both English and French use a vast majority of conjunctive adjuncts in finite clauses, the prominence of this pattern is greater in English (97.9%) than in French (90.7%). The standardised residual for the French data is larger than -1.96, which indicates that this pattern is significantly less frequent in French than in English – where the frequency of CAs used in finite clauses does not differ from the expected frequency. In fact, what the corpus results show is that, whereas English allows for very little variation around the dominant (i.e. finite) pattern, with only 2.1% of the CMs used in other clause types (see Table 14), French appears somewhat more flexible regarding the type of clauses in which it uses CAs of contrast. Firstly, French displays twice as many instances of CAs used in non-finite clauses than English (4% vs 1.7%) – a difference which is significant, as indicated by the standardised residuals. Examples of conjunctive adjuncts used in non-finite clauses are provided in (35) to (37).

(35)   Last August the White House informed Riyadh it was going to strike Syria's Assad regime for using nerve gas against rebel areas, and the Saudis were outraged when Mr Obama declined to do so, **instead** <u>using</u> a deal brokered with Russia on Syria's chemical weapons as a platform to begin a rapprochement with Iran, long the Saudi rival for control of the Gulf.

(36)   On s'en réjouira donc en <u>espérant</u> **cependant** que le gouvernement de François Hollande ne se contente pas de ces appels à la raison.

(37)   La sortie de Valls contre la dépénalisation des délits routiers a laissé apparaître quelques tensions, sans <u>susciter</u> **toutefois** le moindre éclat de voix de Taubira.

In addition, as shown in Table 14, the proportion of conjunctive adjuncts used in verbless clauses is also higher in French (3.3%) than in English, where this pattern is hardly ever found (0.4%). Again, the standardised residuals are much greater than 1.96 in both languages, which indicates that this difference, though small, is statistically significant. Examples of CAs used in verbless clauses are provided in (38) and (39).

(38)   The denials were misleading because the press office had itself been misled. Understandable, you might say, but an error on Mrs Blair's part **all the same**.   (Mult-Ed-EN – The Guardian)

(39)   Apparemment, les faits sont sans rapport les uns avec les autres. Pas tout à fait, **cependant**.   (Mult-Ed-FR – Le Monde)

The frequency difference between French and English is consistent with Chuquet and Paillard (1987: 136–137), who noted a generally higher frequency of nominal clauses in French than in English. Likewise, Guillemin-Flescher (1891: 19–20) stressed that verbless sequences standing on their own are rather frequent in French but can rarely be directly transposed into English. Instead, they generally require the addition of a verb phrase in order to make the process more explicit and avoid a break in the progression of discourse.

   In terms of rank status, the shared preference of English and French for CAs of contrast used in main clauses also needs to be qualified to a certain extent. In the same way as French displays a greater diversity of clause types than English, there is a notable difference between the two languages in their propensity to use conjunctive adjuncts of contrast in main clauses, with English displaying 13.5% more instances of this pattern than French. The standardised residuals represented in Figure 17 indicate that this difference is statistically significant. French compensates for this comparatively low frequency of main clauses by using a statistically significantly greater proportion of conjunctive adjuncts than English in (i) hypotactic clauses and (ii) minor clauses. Embedded uses of CMs such as (40) and (41),

on the other hand, do not give rise to any significant difference between the two languages (as the standardised residuals for this category are lower than 1.96 in both English and French). Examples of conjunctive adjuncts used in hypotactic clauses are provided in (42) to (47), where the host hypotactic clauses are under-lined. As can be gleaned from the examples below, the hypotactic clauses host-ing the CMs can take a variety of forms and functions, among which projection clauses (e.g. (42), (46)), adverbial clauses (e.g. (44), (47)), or relative clauses (e.g. (43), (45)). In addition, the conjunctive adjunct can either relate the hypotactic clause to its main clause – as in (45) – or link two hypotactic clauses – as in (43).

(40)   Quite often, the BBC has used licence-payers' money to outbid its independent rivals for films and sports events <u>that would have been shown free-to-air **anyway**</u>.   (Mult-Ed-EN – The Telegraph)

(41)   Qu'il ne soit pas précautionneux, prudent et réticent, <u>qu'il soit **au contraire** enthousiaste, volontaire et lyrique</u>, est déjà un démenti au soupçon que le premier septennat cohabitationniste de Jacques Chirac avait parfois accrédité d'un engagement européen circonspect.
                                                           (Mult-Ed-FR – Le Monde)

(42)   The police say that Brown attacked the officer and was shot in a scuffle. A friend who was with Brown says <u>that, **on the contrary**, the teenager had his hands in the air</u> and was shouting "Don't shoot" when he was killed.
                                                      (Mult-Ed-EN – The Financial Times)

(43)   The party has thrived through unvarnished bluster, <u>which</u> strikes an authentic note in an age of spin, **yet** <u>conceals chaotic policies which could easily be ripped apart</u>.   (Mult-Ed-EN – The Guardian)

(44)   Downing Street claims that Mrs Blair did not "assist" Mr Foster, who helped her buy two flats, in his legal battle against deportation from the UK. That may be true, <u>although it is **still** astonishing behaviour for someone in her position</u>.   (Mult-Ed-EN – The Independent)

(45)   Et il n'a pas reçu un centime du principal financier du cinéma français, <u>qui avait **pourtant** vocation à l'aider</u>.   (Mult-Ed-FR – Le Monde)

(46)   La garde des Sceaux entend les vider. Avec la meilleure volonté du monde, on ne comprend pas pourquoi cette politique-là empêcherait un assassin de récidiver. On a même tendance à penser <u>que le signal de laxisme lancé par la Chancellerie aurait **plutôt** tendance à l'encourager</u>.
                                                           (Mult-Ed-FR – Le Figaro)

(47)   L'ambition du nouveau ministre de l'économie, Thierry Breton, d'atteindre une croissance de 2,5 % cette année a peu de chances de se réaliser, <u>alors que l'institut table **plutôt** sur 2 %</u>.   (Mult-Ed-FR – Le Monde)

The difference in proportion of hypotactic clauses suggests that French seems to accept the use of conjunctive adjuncts of contrast at non-matrix syntactic level (viz. below the main clause) to a greater extent than English. This may be due to a greater tolerance of French towards the accumulation of relations between the members of a single pair of clauses. In (42) and (46), for example, S1 and S2 are related through both projection and contrast. Likewise, in (45), the contrastive relation is combined with a relation of hypotactic elaboration (in SFL terms), where the subclause provides further information on a given constituent of the main clause (viz. the indirect object). Such a difference is consistent with previous findings, which have shown that co-occurrences of relational markers are more frequent in French than in English (e.g. Crible 2018: 122; see below). However, more research is required to gain deeper insights into this syntactic difference, possibly by analysing the subtypes of hypotactic clauses (viz. relative clauses, adverbial clauses, projecting clauses, etc.) that give rise to the greatest difference between the two languages.

Examples of conjunctive adjuncts used in minor clauses are provided in (48) to (51). Both subtypes of minor clauses, viz. independent clauses lacking a main verb – as in (48) and (50) – and hypotactic clauses standing on their own – as in (49) and (51) – occur in both subcorpora. However, independent clauses without a verb are by far the most frequent type of minor clause hosting conjunctive adjuncts, in both languages (about 70% of the occurrences in English, and just below 80% in French).

(48)   That will further discourage Labour from saying anything pro-European before the general election and could embolden British Eurosceptics, and the Eurosceptic press, to press for a more radical agenda of disengagement. Not necessarily a body blow to Labour, but a severe constraint **nevertheless**.
(Mult-Ed-EN – The Guardian)

(49)   To the rational majority, climate change is not an article of faith but an unwelcome statement of fact. Not because of the anecdotal evidence of unprecedented "weather events" – Australian heatwaves, Philippines typhoons, European flooding – battering the globe with steadily increasing regularity. **Rather**, because scientific study after scientific study has proved beyond all reasonable doubt that not only is the Earth's temperature rising, but those alterations are almost certainly the result of human activities.
(Mult-Ed-EN – The Independent)

(50)   Autre point : les contradictions d'une assistante maternelle, affirmant d'abord qu'un des enfants n'avait pu être influencé dans ses accusations par la télé, avant d'admettre, «les joues cramoisies», qu'il a pu la regarder «machinalement» (le Parisien). **En revanche**, sur le malaise du prêtre-ouvrier, aucune citation.   (Mult-Ed-FR – Libération)

(51)   C'est ici l'occasion de rappeler qu'en France rien n'oblige le chef de l'Etat à
        suivre l'avis de son opinion publique. Même dans le cadre d'une ratification.
        Ce qui paraît **néanmoins** plus qu'improbable.   (Mult-Ed-FR – Le Figaro)

A word should also be said about the borderline category of phrasal-clausal conjunctive adjuncts. As explained in Chapter 6, the phrasal-clausal category was created to include a number of CAs whose host segments could not strictly be viewed as clauses, but which nevertheless closely resembled verbless uses of some subordinators. As shown in Table 14, such ambiguous cases were only found in the French subcorpus. In addition, they mostly concerned the adverbial marker *pourtant*: out of 58 cases of phrasal-clausal occurrences in Mult-Ed, there were 56 occurrences of *pourtant*, one of *au contraire*, and one of *néanmoins*. A close look at these phrasal-clausal cases reveals that they seem to broadly correspond to 'quasi-subordinating' uses of these conjunctive adjuncts. Compare, for example, the behaviour of *pourtant* in (52) and (53), with the subordinators introducing verbless clauses in Examples (54) and (55). There is hardly any syntactic difference between the segments introduced by adverbial *pourtant*, and the verbless hypotactic clauses introduced by the subordinators. The analysis of the syntactic features of the CMs thus made it possible to identify a distinctive use of the French marker *pourtant* which, although commonly described as a resolutely adverbial marker, also allows for some usages bordering on subordination.

(52)   L'Inde elle-même, **pourtant** victime du tsunami, y a vu une «merveilleuse
        occasion» de se poser en grande puissance régionale, rivale de la Chine.
                                                              (Mult-Ed-FR – Libération)

(53)   Le modèle déraille, et Guillaume Pepy, **pourtant** un enfant du TGV, finira
        par s'en rendre compte.   (Mult-Ed-FR – Le Monde)

(54)   Lui aussi, **quoique** homme libre et indifférent à l'opinion des autres, ne
        voulait pas handicaper sa présidence en choquant une partie de l'opinion.
                                                              (Mult-Ed-FR – Libération)

(55)   Yulia Tymoshenko, their best chance, **although** a formidable politician,
        does not look that strongly placed.   (Mult-Ed-EN – The Guardian)

One final comment with respect to the syntactic patterning of CAs of contrast pertains to the features of the first segment in the relation. As explained above, the syntactic analyses performed here concentrate on the features of the second segment in the relation, i.e. the one which contains the conjunctive marker. One feature of S1 was recurrent enough to deserve mention, however. In both English and French, I found a notable number of examples in which the CA expressed a contrast between a hypotactic clause – functioning as the first segment in the relation – and its main clause – corresponding to the host clause of the CM. Such

cases are exemplified in (56) to (59) – where the hypotactic clauses functioning as S1 are underlined. This pattern frequently involved a double signalling of the contrastive relation, as in (56) – where *although* and *nevertheless* are combined – or (59) – where the concessive relation is expressed by both *bien que* and *pourtant*. Interestingly, this pattern was manifestly more common in French than in English, which may testify to a greater tendency of French towards redundancy in the explicit signalling of cohesive relations. Support in favour of this hypothesis is found in Crible (2018: 122), who noted a much greater tendency towards co-occurrence for French than for English discourse markers.

(56)   <u>Although examples of such blatant interference in industry are few and far between</u>, the Government **nevertheless** makes its unhelpful presence felt in another way.   (Mult-Ed-EN – The Telegraph)

(57)   <u>If Russia intervenes again in Ukraine, and war is avoided</u>, the world will **still** become a shrunken, soured place.   (Mult-Ed-EN – The Guardian)

(58)   Le score des forces nationalistes aux élections du 28 décembre, <u>loin d'être un raz de marée</u>, est **cependant** un avertissement aux réformateurs dirigeant le pays.   (Mult-Ed-FR – Libération)

(59)   <u>Bien que le racisme soit une affaire de contenu</u>, il n'est **pourtant** pas étranger à la forme du message dans lequel il s'inscrit, à savoir le genre de discours qu'on appelle sketch ou blague.   (Mult-Ed-FR – Libération)

In addition, the data also revealed cases in which S1 took the form of a verbless absolute or appositive construction, as in (60) and (61). Again, this pattern was mostly found in the French data, in line with previous research underlining that French tends to make a 'copious' use of such constructions, which sound rather old-fashioned in English (Hervey & Higgins 1992: 240; Armstrong 2005: 123).

(60)   <u>Favorable à des sanctions européennes contre Téhéran</u>, elle est **en revanche** opposée […] aux sanctions dans le domaine pétrolier et à des mesures financières ciblées contre les dirigeants de la junte de Rangoun.
                    (Mult-Ed-FR – Le Monde)

(61)   <u>Compliqués, gourmands en temps et en énergie</u>, ces faits d'armes industriels ou financiers n'ont **pourtant** constitué qu'un tour de chauffe avant la véritable épreuve de vérité qui attend le ministre de l'Economie.
                    (Mult-Ed-FR – Le Figaro)

In summary, the study of the syntactic patterning of French and English conjunctive adjuncts revealed that, despite a shared preference for finite and main clauses, English and French also exhibit a number of significant differences regarding the type of structures in which they tend to use CAs of contrast. All in all, French was shown to allow for more syntactic flexibility than English: it displayed a greater

degree of variation with respect to the dominant patterns of use of CAs, both in terms of clause types (finite vs others) and ranks (main vs others).

7.3.1.2   *Syntactic patterns of English and French coordinators of contrast*
The syntactic features of the clauses introduced by the coordinators *but* and *mais* are presented in Table 15. As was the case with conjunctive adjuncts, both English and French primarily use coordinators of contrast in finite and main clauses. This dominant use of coordinators of contrast is exemplified in (62) and (63) – where the finite verbs are underlined. Again, the overwhelming predominance of finite and main clauses was to be expected, given that these patterns are the default syntactic structures in both languages.

(62)   An inquiry into both the Qatar decision and the award of the 2018 tournament to Russia (in preference to England) is nearing completion, **but** its findings <u>are</u> for internal consumption and will not be published.
(Mult-Ed-EN – The Telegraph)

(63)   Ali Zeidan est un homme courageux et compétent, **mais** il se <u>trompe</u> de diagnostic.   (Mult-Ed-FR – Le Monde)

Table 15. Syntactic patterning of English and French coordinators of contrast

Syntactic Features of the Clause		English	French
Clause Type	Finite	94.7%	87.3%
	Non-Finite	2.4%	4%
	Verbless	2.9%	8.7%
Rank Status	Main	93.4%	84.1%
	Hypotactic	2.4%	5.1%
	Embedded	3%	3.5%
	Minor	1.2%	7.3%
Total nb. of Occurrences (Raw)		**1650 (100%)**	**1556 (100%)**

Yet, in line with the results obtained in the previous section, the corpus findings point to a number of significant differences between the two languages. Once again, French appears to display a noticeably greater degree of variation around the dominant syntactic patterns than English (viz. 13% for clause types and 16% for ranks, against 5% and 7%, respectively, for English). In addition, chi-square tests of independence returned significant differences between English and French for both clause types ($\chi^2 = 58.4$, $df = 2$, $p < 0.001$, $\varphi = 0.13$) and rank status ($\chi^2 = 96.6$, $df = 3$, $p < 0.001$, $\varphi = 0.17$), even though the standardised residuals

associated with these two tests – provided in Figures 18 and 19 – indicate that the differences between English and French are less marked than for conjunctive adjuncts.
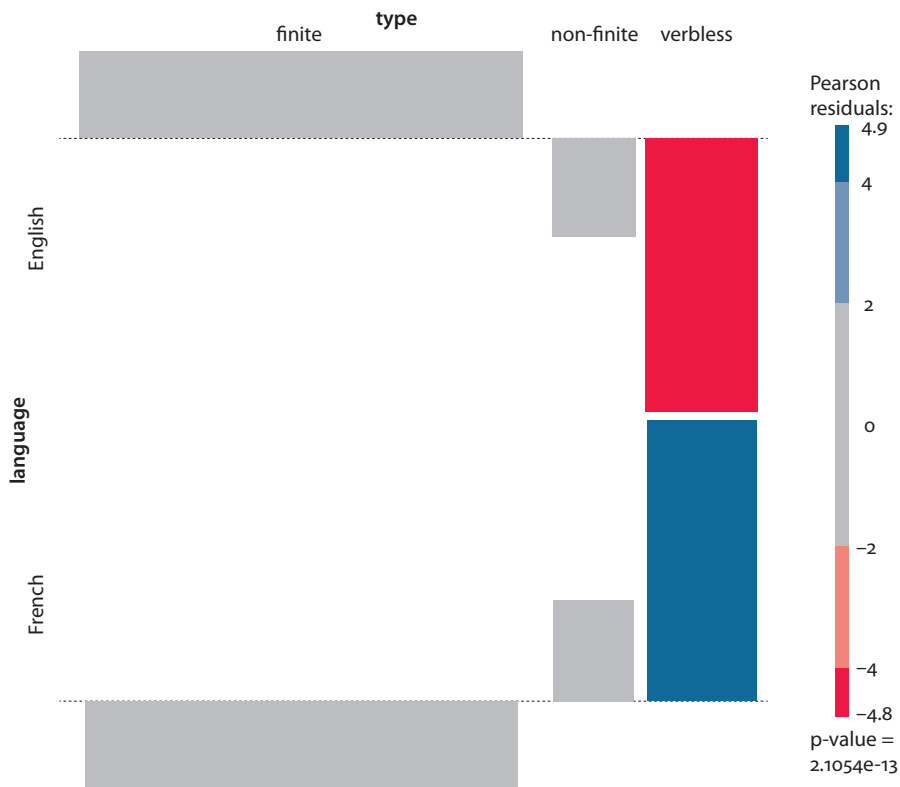
Figure 18. Standardised residuals for the types of clauses hosting English and French coordinators of contrast

Figure 18 shows that in terms of clause types, only the category of verbless clauses gives rise to a significant difference between English and French, with French displaying significantly more coordinators introducing verbless clauses than English – as in Examples (64) to (67). By contrast, the proportions of coordinators used in finite clauses – as in (62) and (63) above – and non-finite – as in (68) and (69) – do not differ significantly across languages.

(64)   The case for greater flexibility is easily advanced, **but** at this juncture much more difficult to attain.   (Mult-Ed-EN – The Telegraph)

(65)   L'alternative est stratégique, **mais** le dilemme très personnel.
                                                                          (Mult-Ed-FR – Libération)
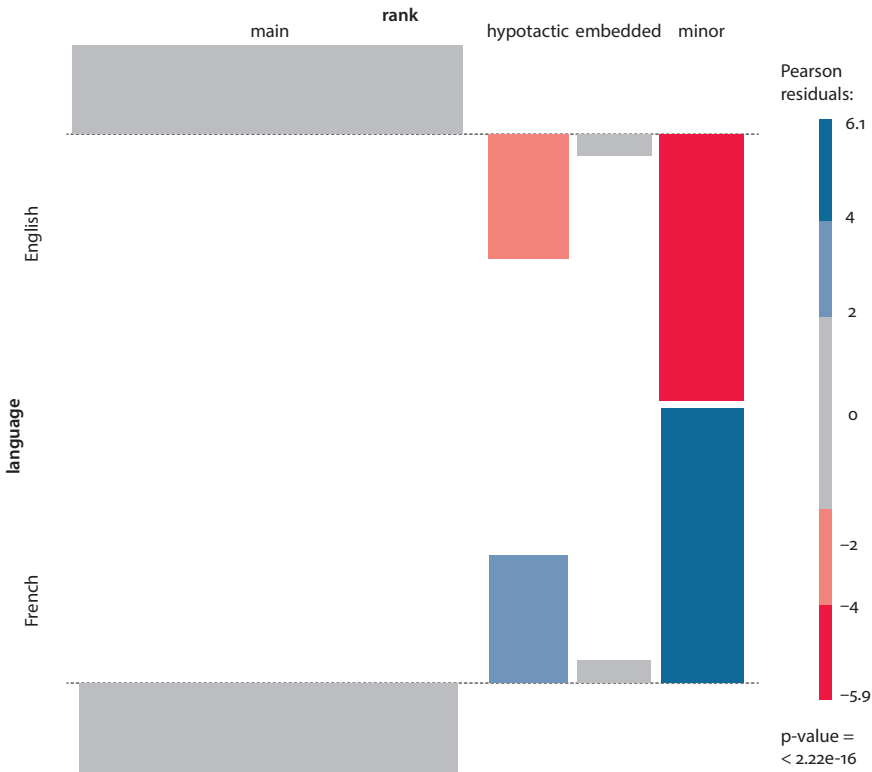
Figure 19. Standardised residuals for the rank status of clauses hosting English and French coordinators of contrast

(66)   After the first day of the first Test, the experts condemned England's spiritless performance, bewailed their injuries, praised the power of the Australians and wrote off the Ashes. **But** not so fast.

(Mult-Ed-EN – The Times)

(67)   Rappelons dans cette perspective que le recours au Conseil de sécurité de l'ONU n'est pas une précaution, mais la garantie du respect du droit international. Non pas un artifice, **mais** un élément majeur de stratégie politique.   (Mult-Ed-FR – Le Figaro)

(68)   European regulators of the single market should strive not to "harmonise" taxes and social policy across the continent, **but** to foster growth on the principle of comparative advantage among differentiated economies.

(Mult-Ed-EN – The Telegraph)

(69)   L'exercice doit être concis, **mais** comment l'être quand on doit dresser le bilan de 2011 et évoquer les perspectives de 2012?

(Mult-Ed-FR – Le Figaro)

It is striking that, in both English and French, the proportion of coordinators used in verbless clauses is noticeably higher than for conjunctive adjuncts, with 2.9% of English and 8.7% of French coordinators used in these structures (as opposed to 0.4% and 3.3%, respectively, for conjunctive adjuncts). This difference is partly due to cases of verbal ellipsis in the second part of a paratactic clause complex – as in (64) and (65) – especially in English, where this pattern accounts for more than two thirds of the verbless clauses. With conjunctive adjuncts, on the other hand, nearly all the verbless clauses were used on their own, as minor clauses – a syntactic pattern which usually sounds more marked, particularly in English. Examples of coordinators introducing verbless clauses standing on their own are provided in (66) and (67).

With respect to syntactic ranks, Figure 19 sheds light on two significant differences between English and French. Firstly, hypotactic uses of coordinators – exemplified in (70) and (71) – appear to be significantly more common in French than English. Thus, French once again seems to display a greater ease than English to use CMs of contrast at non-matrix level. Similarly, coordinators used in minor clauses – as in (72) and (73) – are shown to be significantly more typical of French than English, where they occur no less than six times less frequently than their French counterparts. Moreover, it is significant that most of the English coordinators used in minor clauses are in fact found in headlines. This is not the case in French, where this structure is very frequently found in the bodies of the texts. These two differences echo the results obtained for the category of conjunctive adjuncts, where hypotactic and minor clauses were also found to be more frequent in French than in English.

(70)   He and his team mates will bear a serious responsibility with which they will surely cope **but** <u>which they must treat with the greatest of care</u>.

                                        (Mult-Ed-EN – The Telegraph)

(71)   Il soulignait que le journalisme n'était pas maudit ou mauvais en lui-même **mais** <u>que certaines pratiques le pervertissaient</u>.

                                     (Mult-Ed-FR – Le Nouvel Observateur)

(72)   Narrow politics aside, the Observer believes that the GCSE syllabus should reflect, as well as it can, our multicultural diversity. Should Austen, Dickens, Hardy and Shakespeare be part of such a list? Of course they should. **But** adjacent to Zadie Smith, Sarah Waters and Meera Syal.

                                     (Mult-Ed-EN – The Observer)

(73)   George Bush et Vladimir Poutine, alliés de circonstance contre un même ennemi, jouent une partition commune. **Mais** pas dans le même but.

                                     (Mult-Ed-FR – Le Figaro)

However, Figure 19 also shows that, despite a difference of nearly 10% between English and French, the proportion of coordinators used in main clauses – as in (74) and (75) – does not differ significantly across languages. The same is true of coordinators used in embedded clauses which, once again, do not give rise to any significant difference between the two languages. This pattern is exemplified in (76) and (77), where both coordinators introduce a clause functioning as complement.

(74)   From the 3rd Marquess of Salisbury to the former Tory MP George Walden, politicians can enrich debate beyond politics with thoughtful pronouncements. **But** many are reluctant to do so for fear of proving "controversial".   (Mult-Ed-EN – The Times)

(75)   À l'époque des caricatures du Prophète dans la presse danoise, on pouvait arguer que le chaudron bouillait dans des sociétés musulmanes sous le joug de régimes autoritaires. **Mais**, aujourd'hui, la donne a changé presque partout.   (Mult-Ed-FR – Le Figaro)

(76)   The solution is not to retreat into more central controls, **but** <u>to move in the opposite direction, making it clear where accountability lies</u>.
                          (Mult-Ed-EN – The Guardian)

(77)   L'objectif n'est pas de pousser à la confrontation **mais** <u>de mobiliser les opinions pour leur faire prendre conscience de la gravité de la situation</u>.
                          (Mult-Ed-FR – Le Figaro)

Investigating the syntactic patterning of English and French coordinators also makes it possible to identify what proportion of these markers is used to link two clauses in a clause complex, and what percentage of coordinating CMs occur sentence-initially, as a relator between two isolated sentences. In Systemic Functional terms, the former pattern corresponds to a paratactic use of *but* and *mais*, whereas the latter realises the SFL system of conjunction, defined in a strict sense as the system that ensures the connections between discourse units at sentence level or beyond (see Chapter 2). Intrasentential (i.e. paratactic) uses of *but* and *mais* are exemplified in (78) and (79), respectively, whereas intersentential uses of the markers are provided in (80) and (81).

(78)   Yes, the election campaign in Britain has some echoes of debates across the Channel, **but** the campaign here has in the main been resolutely parochial and trivial.   (Mult-Ed-EN – The Guardian)

(79)   Rien de décisif ne se fera au cours de ce semestre, **mais** le temps des propositions est revenu.   (Mult-Ed-FR – Le Monde)

(80)   In an ideal world, a more equitable system has much to recommend it. **But** this is not an ideal world.   (Mult-Ed-EN – The Independent)

(81)   Il n'a pas honoré ses promesses ? Certes. **Mais** il a tenu le coup, il a su faire
        face, il a résisté à l'ouragan financier.

<div align="right">(Mult-Ed-FR – Le Nouvel Observateur)</div>

The proportions of inter- and intrasentential uses of *but* and *mais* in the Mult-Ed corpus are provided in Table 16, with the standardised residuals produced by the chi-square test provided between brackets beside each percentage value. These figures are restricted to the instances of *mais* and *but* whose host segments are main clauses.

Table 16. Inter- and intrasentential uses of English and French coordinators
of contrast in Mult-Ed (in percent)

	ENGLISH	FRENCH
INTERSENTENTIAL COORDINATORS	61% (−1.83)	69.5% (+1.99)
INTRASENTENTIAL COORDINATORS	39% (+2.5)	30.5% (−2.7)
TOTAL (RAW)	1541 (100%)	1309 (100%)

As appears from this table, intersentential uses of coordinators of contrast are noticeably more frequent than intrasentential (or paratactic) uses. This is true in both English and French, although a chi square test reveals that the proportion of use of each pattern differs significantly across languages ($\chi^2 = 20.84$, df = 1, $p < 0.001$, $\varphi = 0.09$): the standardised residuals indicate that intrasentential coordinators are more typical of English, whereas intersentential coordinators are slightly more typical of French (with the standardised residual for French just topping the 1.96 significance threshold).

The very high proportion of sentence-initial coordinators in the corpus may appear fairly surprising at first sight, in view of the prescriptive rules that have long prevailed against the use of coordinators in sentence-initial position, particularly in English (which may explain the significant frequency gap between the two languages). Although, as underlined by Bell (2007: 183) and Cotter (2003: 52), a majority of usage guides have now considerably relaxed their injunctions against sentence-initial uses of coordinators, these prescriptions have remained very vivid in many users' minds.[8] According to Burchfield

---

8.   Zinsser (2001: 74), for instance, claims the following: "[m]any of us were taught that no sentence should begin with 'but'. If that's what you learned, unlearn it" (see also Burchfield 2004: 121; Lynch 2007 for similar statements). With respect to French, Grevisse and Goosse (2011: 1393) explain that, although some purist grammarians have claimed that it is a mistake to use a coordinator after a full stop, authentic usage does not reflect this prescriptive rule.

(2004: 121), for instance, "the widespread public belief that *But* should not be used at the beginning of a sentence seems to be unshakeable". In the same vein, Struck (1965: 42) points out that "[a]lmost all expert writers […] are convinced that it is stylistically immoral to open a sentence with a coordinate conjunction". The corpus results obtained here, however, seem to suggest that English and French editorial writers have integrated the evolution of the stylistic guidelines in this respect fairly well, since well over half of the coordinators in the corpus are used sentence-initially. These findings are in line with Hundt and Mair (1999: 225), who uncover an increase in frequency of sentence-initial coordinators in journalistic prose over the last third of the twentieth century. Likewise, Cotter (2003) reports a sharp rise in frequency of sentence-initial coordinators in American newspapers over the twentieth century. She adds that opinion texts were among the first type of news reports to use sentence-initial coordinators, since this subgenre is characterised by a certain amount of leeway regarding prescriptive rules, at least as compared to more 'regular' news texts (ibid.: 60). A closer, more qualitative scrutiny of the results shows that the disregard for such a longstanding prescriptive rule in newspaper editorials is not coincidental but fulfils rhetorical and stylistic effects at discourse level, which are particularly in line with the communicative goals of the register. These aspects will be discussed in detail in Section 7.3.2.

7.3.1.3　*Syntactic patterns of English and French subordinators of contrast*
Table 17 summarises the syntactic features of the clauses introduced by English and French subordinators of contrast in the Mult-Ed corpus. As was the case for both conjunctive adjuncts and coordinators, English and French subordinators display the same unmarked syntactic patterns of use: a great majority of subordinators of contrast are included in finite and, unsurprisingly, hypotactic structures, as illustrated in Examples (82) and (83).

(82)　**While** there is no guarantee that sanctions would end Russia's expansionism, they would increase the pressure on Moscow to retreat.
(Mult-Ed-EN – The Financial Times)

(83)　Au début des années 1920, l'Ossétie du Nord a été rattachée à la Fédération de Russie, **alors que** l'Ossétie du Sud a été donnée à la Géorgie.
(Mult-Ed-FR – Le Monde)

Yet, still in line with the results obtained for the other two categories of CMs, English and French also exhibit significant differences in the types of structures in which they use subordinators of contrast. The comparison of both clause types and rank status once again gave statistically significant results (clause types: $\chi^2 = 114.3$, df = 2, $p < 0.001$, $\varphi = 0.15$; rank status: $\chi^2 = 89$, df = 1, $p < 0.001$,

Table 17. Syntactic patterning of English and French subordinators of contrast

Syntactic Features of the host Clause		English	French
Clause Type	Finite	85.6%	92.8%
	Non-Finite	7%	6.1%
	Verbless	7.4%	1.1%
Rank Status	Main	/	/
	Hypotactic	99.6%	95.8%
	Embedded	0.03%	0%
	Minor	0.4%	4.2%
Total nb. of Occurrences (Raw)		**2917 (100%)**	**2210 (100%)**

$\varphi = 0.13$).[9] The standardised residuals for these two categories are provided in Figures 20 and 21.

In terms of clause types, the corpus results show that, as opposed to the findings obtained for CAs and coordinators, it is English which appears to display the more flexibility with respect to the dominant finite pattern: it allows for twice as much variation around the finite pattern than French (viz. 14% vs 7%, see Table 17). Figure 20 indeed shows that with respect to subordinators of contrast, finite clauses are significantly more frequent in French than in English. By contrast, English displays a significantly greater propensity than French to use subordinators in verbless clauses, as in (84) and (85). This result again stands in contrast with the results obtained for the other two categories, where French displayed higher frequencies of verbless clauses than English. Subordinators introducing non-finite clauses as in (86) or (87), on the other hand, are used with roughly similar frequencies in the two languages – and thus appear in grey on the association plot.

(84)   As the cemeteries in Normandy testify, it was a costly battle, **though** mercifully less costly than the Western Front in 1914–18.

(Mult-Ed-EN – The Telegraph)

(85)   **Quoique** l'une des plus passionnantes au monde, la vie de patron de presse est le contraire d'une sinécure.   (Mult-Ed-FR – Libération)

---

9.   As the corpus only contained a single occurrence of subordinators in embedded clauses, this data point was excluded for the computation of the chi-square value. This was necessary to ensure the validity of the statistical results, since a crucial assumption behind the chi-square test of independence is that expected frequencies should exceed five in all the cells of the table (see e.g. Field et al. 2012: 818).
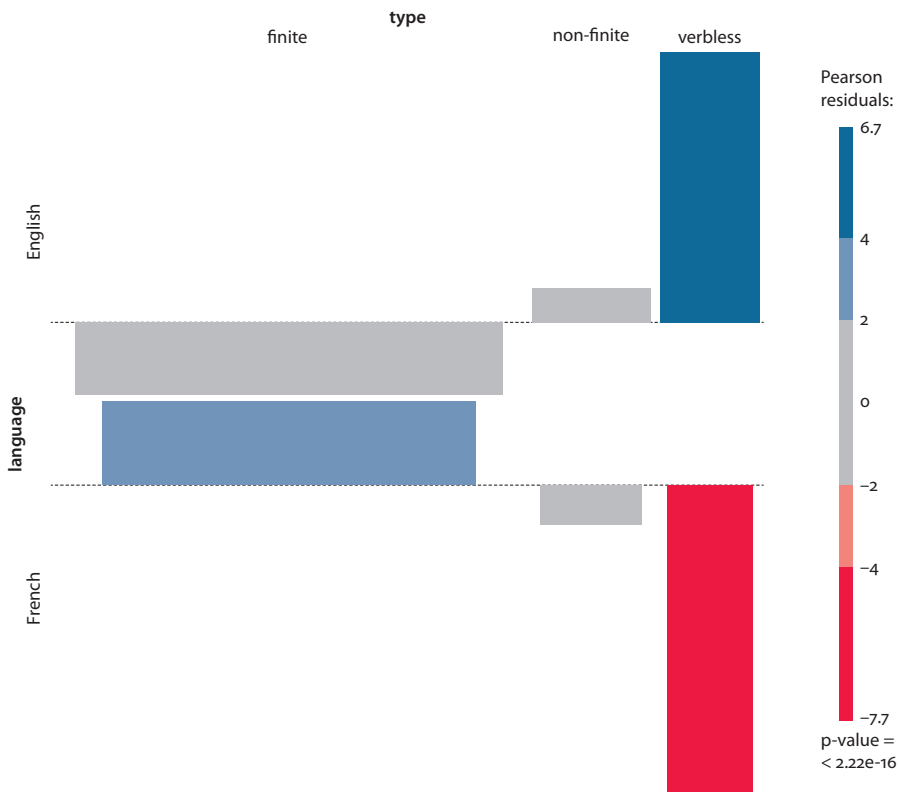
Figure 20. Standardised residuals for the types of clauses hosting English and French subordinators of contrast

(86)   **Although** <u>condemned</u> by Sikh elders, some younger Sikhs spoke up in favour of the play.   (Mult-Ed-EN – The Guardian)

(87)   Israël, **bien qu'**<u>ayant</u> implicitement accepté les termes de cette résolution par le fait même de son admission à l'ONU, ne les a jamais reconnus de facto.
                                                         (Mult-Ed-FR – Libération)

Interestingly, the differences between English and French preferred types of host clauses for subordinators stand in direct contradiction with some of the statements found in the intuition-based contrastive literature. Salkoff (1999: 170), for example, states that "[m]any French subordinate clauses can be shortened by the application of zeroings that are not possible in English". For Salkoff (ibid.), translating such structures into English necessitates the addition of both a subject and a verb, as exemplified in (88). Salkoff goes on to explain that "[a] similar problem arises with those types of clauses taking the form of a subordinate conjunction followed by an infinitive string", as in (89) – although this structure did not give rise
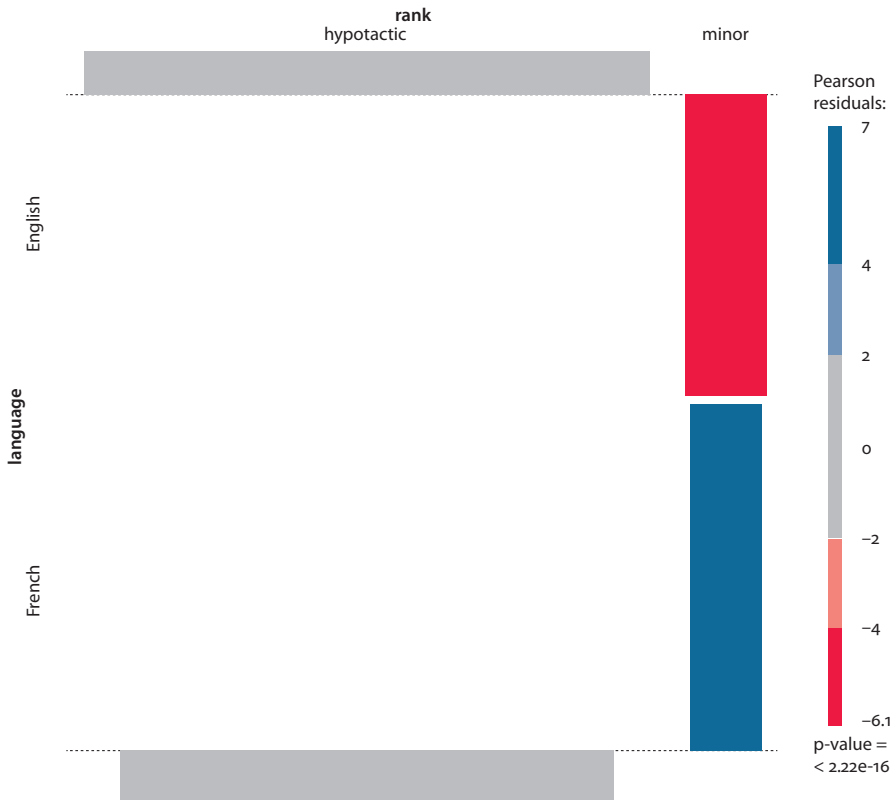
Figure 21. Standardised residuals for the rank status of clauses hosting English and French subordinators of contrast

to any cross-linguistic difference in the present corpus study. Similar patterns of correspondence are underlined by Hervey and Higgins (1992: 231), although they acknowledge that the addition of a verb in translation into English is less necessary with the subordinators *while* and *although* than with a marker such as *because.*

(88)   FR. Quoique discrète, elle dérange les enfants
EN. Although she is discreet, she disturbs the children.   (Salkoff 1999: 203)

(89)   FR. Faute d'y être allé, Max a raté un beau spectacle.
EN. Because he hadn't gone there, Max missed a beautiful show. (ibid.: 171)

With respect to rank status, Table 17 shows that both English and French show very little (viz. less than 5%) variation with respect to the dominant, hypotactic pattern. In spite of this, Figure 21 indicates that French displays a significantly higher proportion of minor clauses than English – a result which is in perfect accordance with those obtained for both coordinators and conjunctive adjuncts.

In both languages, minor clauses introduced by a subordinator all correspond to cases of hypotactic clauses standing on their own without being attached to any main clause, as exemplified in (90) and (91).

(90) When it is said of war, "never forget", that surely means we should remember the sacrifices of the present as well as those of the past. **Although**, as the centenary of the outbreak of the First World War approaches, people's minds inevitably return to history.

(Mult-Ed-EN – The Telegraph)

(91) S'il ne s'agit que «d'une association de queutards», comme l'a élégamment qualifié un policier lillois, rien à dire, ni à redire. **Même si** on peut se demander si l'échangisme est bien le stade suprême du socialisme.

(Mult-Ed-FR – Libération)

In summary, this section has provided a descriptive overview of the syntactic features of the clauses introduced by English and French conjunctive markers of contrast, underlining a number of significant differences between the two languages. Admittedly, the differences underlined in this discussion mostly pertain to the least frequent syntactic categories (e.g. verbless clauses only account for 0.4% and 3.3% of the English and French CAs, respectively). In this, the corpus findings echo Schmied's (2009: 1131) general assertion on contrastive linguistics, viz. that "in typologically similar languages […] contrasts are often not clear-cut but gradient and a matter of choice between options". Faced with such small differences, one may wonder whether it was really worth spending so much time and effort on analysing syntactic features which, for the most part, turned out to be broadly similar across languages. I would like to argue that the differences underlined here, although they are situated at a very high level of granularity, are nevertheless worthy of discussion. Firstly, despite the fact that the patterns differentiating the two languages only account for a minority of the corpus examples, they are among the elements that will contribute to the naturalness of a translation, or a learner's production. The unmarked patterns of English and French, which have been demonstrated to be similar, are surely unlikely to give rise to translation or learning problems. However, the exact transposition of a contrastive sequence made up of two verbless clauses from French into English, or an overreliance on conjunctive adjuncts used in English hypotactic clauses, may very well result in a clumsy-sounding cohesive structure. It is therefore useful to be aware of the patterns which can be transferred directly from one language to the other, and those which run the risk of making the text sound inauthentic – even if these very patterns are not the most frequent ones (see also e.g. Chuquet & Paillard 1987: 135 on the importance of abiding by the syntactic preferences of the target language to ensure the naturalness of a translation). Secondly, and perhaps even more importantly,

the various syntactic uses of conjunctive markers of contrast do not appear to be equivalent: in some cases, using a given CM in a verbless instead of a finite clause, or a minor rather than a main clause, has the potential of creating a range of rhetorical and stylistic effects at discourse level. Consequently, investigating such patterns in a contrastive perspective makes it possible to gain valuable insights into the strategies of textual development of the languages compared. Such issues are examined in the following section.

7.3.2   The syntax-discourse interface

A number of studies over the past decades have been concerned with the description of linguistic phenomena straddling the syntactic and the discourse levels of the linguistic system. Some researchers have highlighted that certain areas of language need to be approached from the angles of both syntax and discourse in order to be described accurately, thus being situated at the syntax-discourse interface (sometimes also referred to as 'grammar-discourse' or 'syntax-pragmatics' interface). On the one hand, some researchers have shown that a range of choices made at the syntactic level are conditioned by the discourse features of the surrounding linguistic context. Travis and Cacoullos (2012: 665–666), for example, show that the choice to express the optional first-person *yo* in Spanish partly depends on a number of discourse-related factors, including subject continuity (i.e. is the first-person pronoun co-referential with the immediately preceding subject), or consistency with the previous realisations in text (viz. *yo* is more likely to be used if the preceding co-referential subject was also expressed).

On the other hand, some linguists have demonstrated the influence of syntax on discourse phenomena. In the field of cohesion, for example, Le Draoulec and Bras (2006) have shown that depending on its syntactic position in the sentence, the French marker *alors* performs distinct functions. When it is used initially, *alors* functions as a conjunctive marker, whereas in medial position, *alors* can only function as a temporal adjunct. In other words, for Le Draoulec and Bras (2006), initial position is a prerequisite for *alors* to perform a function in discourse. Similarly, in a diachronic perspective, Degand and Fagard (2011) have demonstrated that the functional evolution of *alors* seems to go hand in hand with the positional diversification of the adverb. Whereas in Old French, *alors* was predominantly medial, and mainly expressed relations of temporality, over time the marker started being used both initially and medially. In parallel, *alors* developed conditional, causal and metadiscursive functions in addition to its initial temporal meaning. In a diachronic corpus of spoken and written texts ranging from the 12th to the 20th century, Degand and Fagard (ibid.: 51) found a significant correlation between the syntactic position of *alors* in the sentence and its function in discourse: in Middle,

Classical and, especially Present-Day French, temporal uses tend to be associated with medial positions, whereas the other uses of *alors* usually occur at sentence periphery. Starting from the observation that the syntactic evolution of *alors* generally precedes the functional changes, Degand and Fagard (ibid.: 49) conclude that "the variety of syntactic positions opens up possibilities for semantic meaning change". For them, the evolution of *alors* is therefore "illustrative of the way sentence grammar and discourse grammar interact, [with] changes at the sentence level lead[ing] to changes at the discourse level" (ibid.: 29–30).

In fact, the syntax-discourse interface is probably best described as an area of language where syntax and discourse exert a mutual influence on each other. One very clear example of this reciprocal influence pertains to the study of clefting, dislocation and other extraposed syntactic structures. A large number of studies have demonstrated that such syntactic constructions have the potential of achieving a number of rhetorical or stylistic effects. For example, numerous studies have shown that left dislocation – defined as a syntactic structure in which "some [noun phrase] appears in pre-clausal position, coreferential with a personal pronoun occurring somewhere else in the clause" (Prince 1997: 119) – may trigger a range of rhetorical functions at discourse level. These include (i) highlighting a contrast with a preceding segment – as in (92), where *you* is contrasted with *your friends*; or (ii) laying emphasis on a new (or at least non-salient) referent in discourse to facilitate processing by the reader/listener – as in (93), where the appearance of a new participant in the event described (viz. the landlady) is brought into focus through syntactic dislocation (see e.g. Ziv 1994; Prince 1997; Erteschik-Shir 2007; Tizón-Couto 2012).

(92)   I can let you in, but your friends, they must just stop where they are.

(Tizón-Couto 2012: 347)

(93)   My sister got stabbed, she died. Two of my sisters were living together on 18th Street. They had gone to bed, and this man, their girlfriend's husband, came in. He started fussing with my sister and she started to scream. The landlady, she went up, and he laid her out.   (Prince 1997: 122)

However, one other way of looking at the issue is to consider that it is not so much the syntactic structure which triggers a given discourse effect, as it is the pragmatic circumstances that call for specific syntactic structures. This is the view taken by Lambrecht (1988: 138), who claims that "grammatical form is *in part* determined by the pragmatic circumstances under which the sentence is used as a unit of information. In particular, […] syntactic constructions are often *motivated* by various pragmatic requirements having to do with the processing of information in discourse" (ibid.). Such cases thus really represent a kind of 'chicken-and-egg' situation, where syntax and discourse may be said to exert a reciprocal influence on each other.

In this section, I attempt to demonstrate that the syntactic patterning of French and English conjunctive markers of contrast described in the previous section is partly a manifestation of the syntax-discourse interface. I would like to argue that the use of conjunctive markers of contrast in non-default (viz. other than finite and/or main – or hypotactic, for subordinators) syntactic structures is often motivated by and/or creates a range of effects at the discourse level. As a result, the cross-linguistic differences in syntactic patterning of English and French CMs of contrast are shown to be revealing of partly distinct strategies of textual development in the two languages. Importantly, discourse is here defined in a rather broad sense: it encompasses a wide range of rhetorical and stylistic effects which are triggered by the communicative situation in which a given text is produced (viz. newspaper editorial writing; see e.g. Biber & Gray 2010: 15; Staples et al. 2015: 505 for similar definitions of discourse).

7.3.2.1 *Syntactic fragmentation as an emphatic device*
Among the usage patterns that were uncovered through the syntactic analysis of English and French conjunctive markers of contrast, a range of structures give the impression of a somewhat fragmented style of expression. The idea of fragmentation in discourse was first put forward by Chafe (1982: 38), and broadly refers to "the segmentation of information into small, syntactically simple chunks of language that present roughly one idea at a time" (Schiffrin 2006: 186). More specifically, in the Mult-Ed corpus, three main types of syntactic patterns appear to create fragmented textual effects, viz. (i) coordinators and conjunctive adjuncts used in minor clauses – as in Examples (94) to (97); (ii) hypotactic clauses standing alone – as in (98) and (99) ; and (iii) sentence-initial coordinators – as in (100) and (101).

(94) We in Britain, and especially Northern Ireland, are reasonably familiar with the concept, having endured the smaller scale bombing campaign of the IRA in the two decades before 1996. Not as familiar as the Israelis, **however**, who already knew that they were a nation under siege.
(Mult-Ed-EN – The Independent)

(95) Réalisé par Abdellatif Kechiche, l'Esquive sortait mercredi, accompagné de la bonne réputation de son auteur et d'une critique partout louangeuse (Libération du 6 janvier) ; à son détriment, **cependant**, l'affiche du distributeur qui le promeut, et dont son réalisateur n'est autrement responsable que de l'avoir peu ou prou agréée.   (Mult-Ed-FR – Libération)

(96) The suspicion must be that Binyamin Netanyahu, the Israeli prime minister, emphasised the recognition issue precisely because he knew it would be unacceptable to the Palestinians while seeming reasonable to many Jews abroad, particularly in the United States. Israel would thus escape blame, both for continued settlement activity and for the failure of Mr Kerry's initiative. Clever, **but** not smart.   (Mult-Ed-EN – The Guardian)

(97)   Maintes fois évoquée, toujours reportée, la réforme est enfin sur le point
       d'être engagée. **Mais** sur de très mauvaises bases.   (Mult-Ed-FR – Le Figaro)

(98)   The referendum, however, is also about something that it is difficult to
       price – namely exercising power in your own name. **Even if** that sort
       of power comes with new levels of responsibility.

                                                   (Mult-Ed-EN – The Observer)

(99)   Les amendements apportés par le couple américano-britannique à leur
       projet de résolution ont été bien accueillis par le Conseil de sécurité de
       l'ONU. **Même si** plusieurs de ses membres considèrent qu'ils demeurent
       encore insuffisants.   (Mult-Ed-FR – Le Monde)

(100) It is true that most people of retirement age will either leave their pensions
       alone or invest them wisely. **But** there are many more who will not.

                                                (Mult-Ed-EN – The Independent)

(101) Il y a une désescalade. Au lieu de recenser 50 à 100 morts par jour,
       l'Observatoire syrien des droits de l'homme (OSDH), qui tient cette sombre
       comptabilité, en dénombre moitié moins. **Mais** le sang n'a jamais cessé de
       couler.   (Mult-Ed-FR – Le Monde)

The structures exemplified above are all characterised by some degree of marked-
ness, at least in writing. Even though they appear to be more common in French,
in both languages main clauses lacking a verb remain atypical or even 'unortho-
dox' in written language (see e.g. Hervey & Higgins 2002: 116; Carvalho 2005: 95).
Likewise, with respect to hypotactic clauses standing on their own, Rossette
(2009: 28) explains that operating a strong typographical separation between con-
stituents which are syntactically dependent on each other tends to strike readers
as being very marked in written language. Finally, as discussed earlier, sentence-
initial uses of coordinators in writing have long been proscribed as stylistically
infelicitous, although the strong prescriptive rules in that respect are now starting
to lose ground (cf. Section 7.3.1.2). Accordingly, it may be hypothesised that such
marked uses of CMs typically result from a (semi-) conscious will, on the part of
writers, to achieve specific discourse effects. In this section, I attempt to show that
the syntactic fragmentation characterising such uses of CMs is indeed function-
ally motivated.

       In the literature, syntactic fragmentation has usually been associated with
spoken language. In his influential article, Chafe (1982: 38–39) opposed the frag-
mented style of spoken language, which proceeds by the juxtaposition of brief,
usually syntactically basic units, to the integrated style of written language, where
information tends to be packaged into complex, tightly-knits units (e.g. through
subordination or nominalisation; see also Altenberg 1984: 48 for a similar

distinction applied specifically to CM usage).[10] In fact, Chafe (1982: 38) explicitly identified both sentence-initial coordinators and hypotactic clauses standing alone as typical examples of fragmented syntax. Accordingly, Cotter (2003), who observed a sharp rise in frequency of sentence-initial coordinators in American newspapers over the twentieth century, interpreted this change as a reflection of the evolution of journalism as a whole towards a more speech-like, interactional and reader-oriented style of expression. Likewise, Hundt and Mair (1999: 225), who uncovered an increase in frequency of sentence-initial coordinators in journalistic prose over the last third of the twentieth century, also traced this phenomenon back to an evolution of the genre towards a more involved or oral style (see also Chang & Swales 1999: 150; Hyland & Jiang 2017: 44 on the informal character of sentence-initial coordinators). Interestingly, this is consistent with the relatively common use of the conjunctive adjuncts *though, still* and *anyway* in the English subpart of Mult-Ed, which was also postulated to be due to the progressive 'colloquialisation' of newspaper language (see Section 7.2.3.2). The increasingly informal or speech-like nature of newspaper writing most probably explains the strikingly frequent use of sentence-initial coordinators in the corpus (viz. c. 60% in English and 70% in French, see Section 7.3.1.2), along with the use of CMs in minor clauses, which also contributes to creating a partly fragmented textual development. In addition, however, I would like to argue that such syntactic uses of CMs may also be accounted for by reference to the rhetorical effects that they may achieve. More precisely, I intend to show that the use of conjunctive markers of contrast in fragmented syntactic structures can serve emphatic discourse functions – an interpretation that is of course by no means incompatible with the speech-like features of these syntactic patterns.

Sentence-initial uses of coordinators, for example, have sometimes been claimed to increase the force of the cohesive link, by bringing special emphasis to both the logico-semantic relation expressed by the conjunctive marker, and the segment introduced by the CM (Struck 1965: 44; Smith & Frawley 1983: 359; Bell 2007: 196). In other words, as noted by Rudolph (1996: 288), sentence-initial coordinators of contrast give more expressive force to the contrastive relation, thereby contributing to building more powerful argumentative developments. This phenomenon is clearly at play in the Mult-Ed corpus. In Example (102), for instance, the typographical isolation of a very short (i.e. three-word) segment

---

10. Note that the lack of syntactic complexity of speech was later called into question by Biber and Gray (2010), who demonstrated that in some ways, speech is more syntactically elaborated than writing.

through the sentence-initial use of *but* gives special argumentative weight to S2, and emphasises the unexpected nature of its content. Likewise, in (103), the use of two separate sentences instead of a single paratactic clause complex emphasises the contrast between what the writer was hoping for, and what really happened. An intrasentential use of *mais* in the very same example, as in (104), would have reduced the argumentative sharpness of S2 perceptibly. Finally, Example (105) comes from an article written as a reaction to a campaign encouraging young citizens not to vote, as a way to manifest their rejection of the establishment. In this example, the sentence-initial use of the coordinator serves to maximise the impact of a segment whose mere content is already rather strong, viz. that not voting is simply stupid. By being typographically isolated, this sentence is assigned a particularly 'punchy' argumentative value.

(102) The memory of a great nation divided black against white as a matter of racial theory and institutionalised repression appears almost absurd. How on earth could apartheid endure for so long? **But** it did.

(Mult-Ed-EN – The Observer)

(103) Comment des adultes peuvent-ils se permettre de manipuler ainsi une enfant ? On aurait aimé que cet incident soit une exception. **Mais** c'est tout le contraire qui se révèle.   (Mult-Ed-FR – Libération)

(104) Comment des adultes peuvent-ils se permettre de manipuler ainsi une enfant ? On aurait aimé que cet incident soit une exception, **mais** c'est tout le contraire qui se révèle.   (adapted version of (103))

(105) Young people are supposed to be angry. They tend to feel injustices more keenly than their elders and to be more impatient to put them right. **But** they do not have to be stupid.   (Mult-Ed-EN – The Independent)

The rhetorical power afforded by the use of punctuation is also noted by Hervey and Higgins (2002: 116–117). According to them, segmenting information units by means of strong punctuation marks to achieve emphasis is a technique which is commonly found in persuasive registers such as journalism or advertising. Among other phenomena, Hervey and Higgins (ibid.: 116) mention the case of unorthodox punctuation separating a hypotactic clause from its matrix clauses, also insisting that such structures grant focus to both the isolated dependent clause, and the cohesive link that introduces it. Such a strategy appears clearly from the analysis of the hypotactic clauses standing alone in Mult-Ed. In Examples (106) and (107), for instance, the segmentation of the clause complex into two clearly distinct parts serves to emphasise the contrast expressed by the subordinator, as though the writer was seeking to stress the absurdity of what is stated in S1. The rhetorical effect achieved by punctuation in these examples appears very clearly when comparing (107) with its non-fragmented equivalent in (108). In (108), the

subclause introduced by *alors que* is granted far less focus, and consequently loses some of the argumentative weight carried by the equivalent discourse segment in (107). It is also noteworthy that in other cases, such as (109) and (110), the isolated hypotactic clause gives the impression of an afterthought, which the writer adds to somehow qualify the content of the preceding segment. Importantly, the use of hypotactic clauses standing alone in discourse had already been noted by Altenberg (1984: 48–49) in his study of causal linking in English speech and writing. Altenberg also interpreted such uses as signs of fragmentation in discourse, and found them to be especially common in spoken language – which constitutes yet further evidence of an evolution of journalese writing towards more informal stylistic conventions.

(106) Aussi bien Ennahda, en Tunisie, que les Frères musulmans, en Egypte, entendent imposer leur ordre coranique, par la force si nécessaire, à leurs concitoyens. **Alors que** ni l'un ni l'autre de ces partis, vainqueurs certes d'élections libres dans des pays qui n'avaient connu que des dictatures, n'est majoritaire. (Mult-Ed-FR – Libération)

(107) Les commentateurs se focaliseront sur la beauté des mouvements diplomatiques et l'habileté des compromis. **Alors que** tout se jouera, aujourd'hui et demain, dans la gestion des conséquences sociales de la crise. (Mult-Ed-FR – Libération)

(108) Les commentateurs se focaliseront sur la beauté des mouvements diplomatiques et l'habileté des compromis, **alors que** tout se jouera, aujourd'hui et demain, dans la gestion des conséquences sociales de la crise. (adapted version of (107))

(109) The referendum, however, is also about something that it is difficult to price – namely exercising power in your own name. **Even if** that sort of power comes with new levels of responsibility.

(Mult-Ed-EN – The Observer)

(110) Un dialogue s'engage. **Même si** Alain Juppé s'excuse de ne pouvoir répondre à tout le monde. (Mult-Ed-FR – Libération)

Finally, the use of CAs and coordinators in minor clauses also testifies to the emphatic stylistic effects that can be achieved by the syntactic features of S2. Similarly to sentence-initial coordinators and isolated hypotactic clauses, CMs used in minor clauses contribute to enhancing the persuasive tone that is central to editorial writing. In such cases, the typographical isolation of S2 is compounded by the syntactically reduced character of the segment in order to increase the punchy nature of the argumentation (see also Chuquet & Paillard 1987: 137 on the stylistic function of minor clauses). In (111), for instance, the isolated verbless clause lays particular emphasis on the fact that developing countries have actually received

nothing at all. The syntactic form taken by S2 gives this argument a particularly incisive and forceful quality, while also emphasising the contrast with what other countries, such as Australia or New Zealand, have received. Again, the argumentative effect achieved by the use of a verbless S2 appears clearly when comparing (111) with (112), in which a verb phrase was restored in the second segment. The exact same phenomenon is visible in (113), where the use of a verbless minor clause with the coordinator *mais* strengthens the contrast between the amount of attention that was granted to Dieudonné and the Islamic veil, as opposed to Van Gogh's murder, in a way that also highlights both the injustice and the absurdity of the situation. The English example in (114) achieves a comparable effect, with the use of a minor clause increasing the salience of the contrast between the progress made in the fields of racial equality and policing. Through the use of a minor clause, the editorial writer also seems to grant more argumentative force to S2 than S1: in (114), the information that is granted the more focus is the lack of progress affecting the field of policing. It is also worth noting that, similarly to both sentence-initial coordinators and hypotactic clauses used on their own, minor clauses lacking a verb are typically associated with oral language: as Carvalho (2005: 95) explains, "any utterance with this structure, when it appears in writing, has more or less strong oral connotations" [my translation].

(111)   Le faux procès que l'Australie fait à la PAC oublie de mentionner que les parts de marché abandonnées par l'Union européenne depuis 1992 ont été gagnées par des pays comme celui de M. Vaile: l'Australie par exemple a pris la place des producteurs européens dans le domaine des céréales. On peut dire la même chose de la Nouvelle-Zélande pour les produits laitiers. Rien **en revanche** pour les pays en développement les plus pauvres.

(Mult-Ed-FR – Le Figaro)

(112)   Le faux procès que l'Australie fait à la PAC oublie de mentionner que les parts de marché abandonnées par l'Union européenne depuis 1992 ont été gagnées par des pays comme celui de M. Vaile: l'Australie par exemple a pris la place des producteurs européens dans le domaine des céréales. On peut dire la même chose de la Nouvelle-Zélande pour les produits laitiers. Rien n'a **en revanche** été octroyé aux pays en développement les plus pauvres.

(adapted version of (111))

(113)   Nous nous enflammons pour ou contre Dieudonné, pour ou contre le voile islamique, demain pour ou contre les coups de canif à la loi de 1905. **Mais** sur le meurtre de Van Gogh, qui nous parle exactement de la même chose, pas d'émotion, pas d'indignation, pas d'effroi.   (Mult-Ed-FR – Libération)

(114)   Indubitably – and as the election of a black president testifies – the US has made great strides in the past half-century towards achieving racial equality. Less so, **however**, in the field of policing.   (Mult-Ed-EN – The Independent)

Finally, it is notable that the segmentation of argumentation into clearly separate discourse units, usually containing a single piece of information, may also be interpreted as a manifestation of the reader-friendly concerns of current journalistic writing underlined by Cotter (2003) or Westin (2002: 163): by drawing a clear typographical boundary between the segments related by the markers, editorial writers outline the different stages in the argumentative development, which may create a facilitating effect for the reader (see also Bell 2007: 198).

In summary, the syntactic choices made by writers with respect to conjunctive markers of contrast do not appear to be trivial, but rather seem to perform well-defined functions in discourse. Importantly, as was demonstrated in Section 7.3.1, English and French differ in their frequency of use of the syntactic structures discussed in this section. All three patterns (viz. (i) conjunctive adjuncts and coordinators used in minor clauses; (ii) sentence-initial coordinators; and (iii) hypotactic clauses standing alone) were shown to be significantly more frequent in French than in English. Consequently, the differences in syntactic patterning of English and French conjunctive markers of contrast described in Section 7.3.1 seem to testify to divergences between the more general strategies of textual development of English and French. With respect to CMs of contrast, French makes a more extensive use than English of the options afforded by syntax to grant increased emphasis to discourse relations of contrast.[11] In addition, as all three types of syntactic uses have been shown to be typical of spoken language, it may also be hypothesised that the results testify to a greater tendency of French editorials to make use of speech-like syntactic structures as compared to English editorials. More research is nevertheless required in order to corroborate this assumption.

7.3.2.2 *Syntactic compression of contrastive linking in English and French editorials*

Alongside uses of CMs that contribute to the creation of a fragmented – and thus more emphatic – textual development, the corpus data also displays a number of patterns which testify to a certain degree of syntactic compression and/or integration in the expression of conjunctive relations of contrast. By syntactic compression and integration, I mean ways of expressing contrast where the information in S2 is conveyed in a compact and synthetic way (i.e. in a limited number of words, for instance through the ellipsis of redundant or easily-recoverable constituents),

---

11. Note that the role played by French syntax in the expression of contrastive discourse relations is partly in line with the hypotheses formulated in Section 7.2, where devices such as thematisation and emphatic pronouns were suggested as alternative resources to explicit CMs for the expression of contrast in French.

usually resulting in a situation where the related segments are tightly-knit together (e.g. by sharing a subject or a verb, or simply by being syntactically dependent on each other). More specifically, among the syntactic patterns involved in the present analysis, the ones that have also been demonstrated to contribute to more compressed and integrated styles of expression are non-finite and verbless clauses (see e.g. Quirk et al. 1985: 995–996; Greenbaum 1988: 4; Johansson & Lysvåg 1986: 177–178; Granger 1997: 195) and subordination (Chafe 1982: 44; Lehmann 1988; Matthiessen 2002). As noted by Biber (2003: 170), compression is a central linguistic process in newspaper prose. It is mainly motivated by space-saving considerations, in an age when the 'informational explosion' pressures newspapers to convey information in the most economical way possible (see also Westin 2001: 84). In addition, according to Biber (2003: 170), the tendency towards compression is in line with the central informational purposes of the register.

Whereas in the previous section, French was shown to display consistently higher frequencies of devices contributing to the fragmentation of discourse than English, both languages exhibit patterns that ensure an integrated and compact expression of discourse relations of contrast. However, English and French differ in the types of devices to which they resort to express these relations in compressed ways. On the one hand, as explained in Section 7.3.1, French displays a statistically significantly larger number of (i) CAs used in non-finite clauses; and (ii) coordinators and CAs used in hypotactic clauses.[12] Hypotactic uses of conjunctive adjuncts are illustrated in Examples (115) to (117), which also exemplify non-finite uses of these markers. A hypotactic use of a French coordinator is illustrated in (118).

(115)   Au nom de cette approche, le contact a été renoué avec Damas, **pourtant soupçonné d'œuvrer à la déstabilisation du Liban**.

(Mult-Ed-FR – Le Monde)

(116)   Pour de multiples raisons, mais tout particulièrement du fait du caractère aléatoire et expérimental de ce projet, le Comité consultatif national d'éthique vient d'émettre à ce sujet un avis plutôt mitigé, laissant **toutefois la porte entrouverte à une telle intervention** à condition qu'elle soit réalisée dans un cadre très contrôlé.   (Mult-Ed-FR – Le Figaro)

(117)   La sortie de Valls contre la dépénalisation des délits routiers a laissé apparaître quelques tensions, sans susciter **toutefois** le moindre éclat de voix de Taubira.   (Mult-Ed-FR – Libération)

---

12.   Conjunctive adjuncts and coordinators used in verbless clauses were also among the typically French syntactic patterns, but these usually stand on their own, thus functioning as minor clauses. Therefore, these uses rather belong to the patterns which contribute to fragmentation in editorial writing. (cf. previous section).

(118)   L'Occupation, que l'écrivain n'a pas connue, **mais** <u>dont il a vécu des conséquences et tutoyé des ombres</u>, est remontée roman après roman, génération après génération.   (Mult-Ed-FR – Libération)

In all five examples, the relationship is conveyed in a rather syntactically compact way. In (115), for instance, the finite version of the example would have required the use of at least two more words (*qui était pourtant soupçonné*). The use of *pourtant* in a main clause, on the other hand, would have demanded a full stop, along with a repetition of the subject and the use of a finite verb form preceding the past participle. These two formulations would arguably have sounded some-what heavier, whereas Example (115) conveys the information in a more direct, economical and possibly more effective way (see also Altenberg 1984: 48 on the textual 'compactness' achieved by non-finite uses of CMs). Another excellent example of a contrastive relation expressed in a syntactically compressed manner is provided in (118). This use of the coordinator *mais* allows the writer to express three facts about the subject of the sentence (viz. *l'occupation*) in one go. His/her main point is expressed in the main clause, whereas two other clauses standing in opposition, and providing further details on this subject, are added as coordinated non-defining relative clauses. One non-hypotactic variant of this contrastive rela-tion is provided in (119).

(119)   L'écrivain n'a pas connu l'Occupation, **mais** il en a vécu des conséquences et tutoyé des ombres. Elle est remontée roman après roman, génération après génération   (adapted version of Example (118)).

While this example contains the exact same content as the original formulation, it involves the use of three main clauses instead of one, which leads to a certain dilution of the writer's main point. Moreover, the two examples differ markedly in terms of information structure: the alternative formulation in (119) lays the main focus on the element which was used as subject of the hypotactic clauses in the original example (*l'écrivain*), instead of placing *l'Occupation* at the centre of the assertion (i.e. as the Theme of the entire utterance). In other words, the hypotactic use of the marker in (118) also allows the writer to place the whole contrastive sequence in the background. Interestingly, the use of non-defining relative clauses as a means of compression in (English) journalisitic prose was already noted by Biber (2004: 177):

> [Non-defining relative clauses] are used in newspaper prose as compressed de-vices to pack extra information into relatively few words […] They are typically used to present information that is tangential to the main point, but that might be of interest to some readers […]. In other cases, these modifying clauses present tangential information that provides important background for the interpreta-tion of the main story line".

Whereas French tends to resort to both (i) conjunctive adjuncts in non-finite clauses, and (ii) coordinators and CAs used hypotactically to express relations of contrast in a compressed way, in English this type of stylistic effect tends to be achieved through the use of subordinators of contrast in verbless clauses. In Section 7.3.1.3, it was demonstrated that subordinators of contrast introducing a verbless clause, as exemplified in (120) to (122) below, were significantly more common in English (7.4% – or 215 instances) than in French (1.1% of the subordinators, i.e. 24 instances).

(120)   GM products could be treated on a case-by-case basis, and national bodies rather than the EU could be given responsibility for evaluating them, **albeit** under the auspices of a pan-European food body.

(Mult-Ed-EN – The Independent)

(121)   The Agenda for Change talks, **although** lengthy, have been conducted in a positive spirit.   (Mult-Ed-EN – The Times)

(122)   Efforts at political reform in Saudi Arabia, Egypt, Jordan and Kuwait, **while** welcome, are but partial remedies.   (Mult-Ed-EN – The Guardian)

Numerous researchers have underlined the potential of verbless clauses for syntactic compression. In their overview of English clause types, for example, Quirk et al. (1985: 995) explain that "[b]ecause nonfinite clauses lack tense markers and modal auxiliaries and frequently lack a subject and a subordinating conjunction, they are valuable as a means of syntactic compression". They then add that "[v]erbless clauses take syntactic compression one stage further than nonfinite clauses" (ibid.: 996) since, in addition to tense markers, subjects and modal auxiliaries, they also lack a verb (see also Greenbaum 1988: 4). Therefore, as underlined by Johansson and Lysvåg (1986: 178), the main motivation for using verbless clauses is usually to achieve greater informational density. In addition to representing compact ways of expressing contrast, verbless clauses also display an especially high degree of syntactic integration: clauses lacking a finite verb form have usually been said to be more tightly integrated to their surrounding context than their finite equivalents (see e.g. Matthiessen 2002: 174). The increased degree of compression and informational density achieved by verbless clauses is clearly visible in Examples (120) to (122). In (120), for instance, the use of a verbless clause makes it possible to avoid the repetition of both the subject and the verb of the clause which, in this case, constitute quite a long sequence (i.e. *national bodies rather than the EU could be given responsibility for evaluating them*). The same phenomenon is at play in (121) and (122), where the subject and the verb (*they are*, in both cases) are elipted. Admittedly, the subject-verb sequences in these examples are limited to two short words. However, the compressed forms still result in more direct

formulations. Compare, for example, (121) with its 'full', somewhat heavier version in (123). Similarly to the non-defining relative clauses discussed above, these reduced syntactic clauses seem to be used when the reader wants to incidentally present information that is "tangential to the main point" (see Biber 2004: 177), but still relevant for the argumentative development.

(123)   The Agenda for Change talks, **although** <u>they are</u> lengthy, have been conducted in a positive spirit   (adapted version of Example (121)).

It is worth noting that one feature of syntactic compression is the fact that the relations between linguistic constituents tend to be more implicit than when their more elaborated equivalents are used. In their corpus-based study of academic writing, for example, Biber and Gray (2010) identify a number of grammatical structures used by researchers to present information in a compact way. They show that all these structures give rise to a high degree of implicitness in contemporary academic prose, where the readers have to infer most of the links between linguistic units (see also Biber 2004: 179 on this). One example that they give pertains to phrasal (as opposed to clausal) modification, as in *heart disease, alcohol consumption, prison officials* or *union assets*. As Biber and Gray explain (ibid.: 12), in such structures there is no grammatical clue to inform the reader about the relationship between the head noun and the premodifying noun, although a very wide array of relationships may hold between two nouns. For example, while a *heart disease* is a *disease <u>located</u> in the heart, alcohol consumption* refers to the <u>*process*</u> *of consuming alcohol*, and *prison officials* are *officials <u>who work</u> in a prison.* Similarly, the CMs displaying the syntactic features discussed in this section may be said to express contrast in a less explicit way than CMs introducing finite and main clauses. With verbless and non-finite structures, for instance, the reader him/herself needs to supply elements such as the subject, the tense, the verb, etc. (e.g. Johansson & Lysvåg 1986: 98).

In summary, both English and French display their own syntactic strategies to express contrast between two clauses in a dense, synthetic way. When comparing these results with those discussed in the previous section, it is striking that the two discourse effects associated with the syntax of contrastive CMs actually seem to pull in opposite directions. Whereas the patterns discussed in the present section were associated with compression and integration, Section 7.3.2.1 presented a number of syntactic uses of CMs contributing to textual fragmentation. As already mentioned above, these two tendencies have usually been described as two opposites, with fragmentation being typical of oral and informal communicative situations, whereas integration and compression tend to characterise written productions, situated towards the more formal end of the cline (see e.g. Chafe 1982: 38–45; Altenberg 1984: 48; Biber 2004: 171; Biber & Gray 2010). Therefore, the syntactic patterns of English and French CMs of contrast embody a tension

that seems to be at play in contemporary newspaper prose. On the one hand, the use of sentence-initial coordinators and minor clauses (with the latter pattern being especially typical of French) is exemplary of the drift of newspapers towards a more oral, informal and reader-friendly writing style. On the other hand, other syntactic features of contrastive CMs – viz. non-finite and/or hypotactic uses of CAs and coordinators in French, and verbless uses of subordinators in English – testify to the tendency of newspaper prose towards linguistic compression. As opposed to the reader-friendly character of more fragmented textual patterns (such as sentence-initial *but* and *mais*), these syntactic features involve a decrease in the degree of explicitness of the relation, forcing readers to infer certain aspects of the sequence (e.g. by restoring ellipted subjects or verbs). Interestingly, a similar tension was identified by Biber (2004: 179–180) in his study of compressed noun-phrase structures in newspaper discourse, leading him to reach the following conclusion: "at the same time that news has been developing more popular oral styles, it has also been innovative in developing literate styles with extreme reliance on compressed noun phrase structures". Likewise, in her corpus-based study of language change in newspaper editorials, Westin (2002: 164) noted that:

> [T]wo conflicting linguistic paradigms are at work in the editorials: the aspiration for informality and the aspiration for information density and lexical specificity, the former probably being the result of an adjustment to a new and broader reading public and the latter the result of an adjustment to the special "house styles" that developed over the years.

In summary, in this section devoted to the syntax-discourse interface, I hope to have demonstrated the relevance of analysing the syntactic features of the segments in which conjunctive markers tend to be included. Although, as stated in the introduction to this section, this approach may seem surprising at first sight, the corpus analyses have demonstrated how closely these syntactic patterns are in fact related to discourse-level strategies. The corpus results have provided a very valuable window onto (i) the methods of textual development which are at play in English and French newspaper editorials (with the two competing tendencies of fragmentation on the one hand, and compression on the other), and (ii) the differences between the two languages in that respect (e.g. French was shown to display a somewhat greater propensity towards fragmentation than English; English and French were demonstrated to differ in the syntactic structures which they tend to prefer to achieve compression in the expression of logico-semantic relations of contrast).

7.3.3   The syntax-lexis interface

So far, the syntactic features of CMs have been discussed in fairly broad terms, taking the three categories of conjunctive adjuncts, coordinators and subordinators as

unitary groups. Yet, it may be that different conjunctive markers, although belonging to the same grammatical category, display distinct syntactic preferences. This section seeks to investigate the lexical variation in syntactic patterning of contrastive CMs within the two grammatical categories of conjunctive adjuncts and subordinators. Coordinators are not discussed here, since this category contains a single CM per language. The aim is to assess (i) whether all the conjunctive markers within a syntactic category tend to behave in similar ways and (ii) whether the syntactic tendencies and cross-linguistic differences underlined in the previous sections are evenly spread across the markers, or whether they are mostly attributable to a few markers displaying idiosyncratic syntactic features. Therefore, whereas the previous section explored the connections of syntax with an upper level of the linguistic system (viz. discourse), the present discussion rather moves downwards, investigating the relationships between syntax and lexis. The decision to search for such patterns of lexical variation responds to an acute awareness, in corpus-based linguistic research, that in addition to displaying some degree of regularity and repetitiveness, "linguistic patterns are [also] highly idiosyncratic in the sense that they tend to be item-specific and unpredictable, thus defying all attempts at capturing them by general abstract rules" (Herbst et al. 2014: 1). In this respect, Hoey's (2005) theory of lexical priming postulates that each lexical item in a language has a well-defined, idiosyncratic set of features in terms of collocation, colligation, semantic and pragmatic associations, stylistic features, position in the sentence, etc. The aim of this section is to identify whether the distinct lexical items within each language system are primed differently according to the syntactic features of their host segments.

The observations made in this section are fully descriptive: despite the fact that the CMs occurring very infrequently (below 5 occurrences per million words) in the corpus were excluded from the analysis, the very low frequency of occurrences per conjunctive marker in some categories (e.g. non-finite, verbless or minor uses of the CMs) made it impossible to apply any statistical measures to determine the degree of significance of the differences observed.[13] In addition, in order to avoid overwhelming the reader with figures, the discussion will be restricted to a few of the most striking examples which illustrate the relevance of taking lexis into account when describing the syntax of English and French CMs. The full tables,

---

13. Note that the initial objective of these analyses was to resort to multifactorial statistical techniques to assess the respective influence of (i) language and (ii) lexis on the corpus results. However, given that the English and French lexical items could not be paired systematically (due to the impossibility of establishing clear one-to-one mappings between CMs across languages), it was impossible to use that type of statistics on my data.

containing the exact number of occurrences of each marker in each syntactic cat-
egory, are provided in Appendix 2.

   Figure 22 provides the breakdown of the clause types identified in the English
subcorpus per subordinator. The raw frequencies of each marker in the corpus are
provided between brackets. For the sake of visual clarity, the percentage values
that do not reach 2.5% are not displayed on the bar plot. As appears clearly from
this graph, English subordinators partly differ in the type of clauses that they typi-
cally introduce. The most obvious example of the influence exerted by lexis on the
features of the host segments is *albeit*, which is used almost exclusively in verb-
less clauses – as in (124). This contrasts sharply with the types of clauses typically
introduced by the other six subordinators, which display between 80 and 100% of
finite uses.

(124)   Dr David Kelly conveyed, **albeit** in a somewhat mangled and self-serving
        form, the concern of some members of the intelligence world about that
        misrepresentation to the BBC reporter Andrew Gilligan.
                                        (Mult-Ed-EN – The Telegraph)

Alongside the strikingly idiosyncratic behaviour of *albeit*, Figure 22 also reveals
some more minor differences between the markers. Firstly, while *even though* and
whereas do not seem to permit verbless segments, such patterns are rather com-
mon with *although* (44 instances), *though* (80 instances), and *while* (46 instances).
Verbless uses of these three markers are exemplified in (125) to (127).

(125)   Ms Rousseff's approval rating, **although** still well ahead of her opponents, is
        sagging.   (Mult-Ed-EN – The Financial Times)
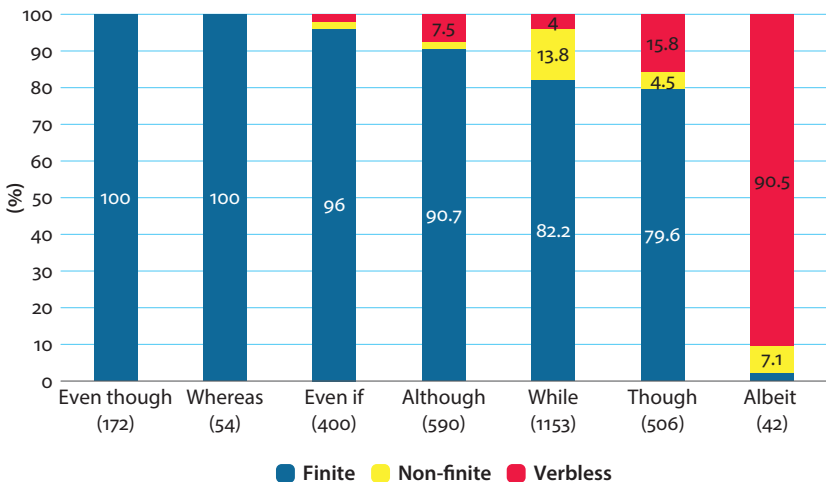

Figure 22. Breakdown of the clause types per subordinator in English (in percent)

(126)   But Crick, **though** a scientist of genius, was not much of a philosopher.
                                                    (Mult-Ed-EN – The Guardian)

(127)   The mansion tax proposed by Labour and the Liberal Democrats, **while** a
         nod in the right direction, is more about symbolism than market correction,
         mainly affecting the highest end of the London market.
                                                    (Mult-Ed-EN – The Observer)

Interestingly, the proportion of verbless *though* is twice as high as that of *although*. Thus, whereas the difference between *though* and *although* has generally been claimed to simply boil down to a divergence in degree of formality of the two markers – with *although* emerging as the more formal variant of the pair (e.g. Quirk et al. 1985: 1097; Biber et al. 1999: 845) – these results suggest that the two subordinators also differ with respect to the syntactic features of the segments which they tend to relate. These partly diverging syntactic preferences justify the separatist approach adopted in the present chapter, where markers that are usually considered as mere variants of a single item are viewed as distinct CMs. Had the occurrences of *although* and *though* been merged, this difference would have gone unnoticed.

The seven English subordinators of contrast also differ in their propensity to introduce non-finite clauses: whereas about one in seven occurrences of *while* are used in non-finite clauses – as in (128) – *even though* and *whereas* are never used in such structures. In addition, out of a total of 400 occurrences of *even if*, only eight introduce a non-finite clause (against 159 instances of *while*, for example). Interestingly, a large majority of the non-finite uses of *while* are not fully contrastive but combine a contrastive meaning with a temporal one. In (128), for instance, the two clauses stand in a relationship of both simultaneity (where *while* could be paraphrased as *at the same time*) and contrast (the two actions performed by Putin are presented as contradictory). This may suggest that there is a relationship between the semantic features of the markers and the syntactic form taken by their host clause, with non-finite clauses being more likely to occur when the subordinator combines a temporal meaning with a contrastive one.

(128)   It was before the World Cup final that Mr Putin met with Angela Merkel
         and assured her that he desired peace in Ukraine, **while** quietly sending yet
         more troops and kit across the border.   (Mult-Ed-EN – The Telegraph)

Figure 23 represents the proportion of use of each French subordinator of contrast in each clause type. The overwhelming majority of finite uses that was already noted in Section 7.3.1.3 is reflected in this lexical breakdown, where six subordinators out of nine are (almost) never used in other clause types. These results thus reaffirm the apparent repulsion of French subordinators of contrast for non-finite and verbless patterns. In fact, only two French markers seem to be acceptable in

verbless clauses, viz. *bien que* and, especially, *quoique* – for which verbless clauses are in fact the most typical host segments. The verbless uses of these two subordinators are exemplified in (129) and (130).

(129)   C'est vrai, dans l'UE, les Roms posent problème. Ils mendient, se regroupent dans des bidonvilles, certains pratiquent toute la palette des conduites délinquantes. Ils gênent et ils inquiètent. Pourtant, ce problème d'insertion, **bien que** très visible, est sans commune mesure avec les problèmes que doivent affronter leur vie durant la très grande majorité de ces Européens oubliés du progress.   (Mult-Ed-FR – Le Monde)

(130)   La possibilité d'une offre par l'italien Finmeccanica, également évoquée, n'est pas forcément intéressante. **Quoique** européen, cet acteur industriel est très éloigné du métier d'Eutelsat et ne constituerait pas un repreneur adéquat.   (Mult-Ed-FR – Le Figaro)
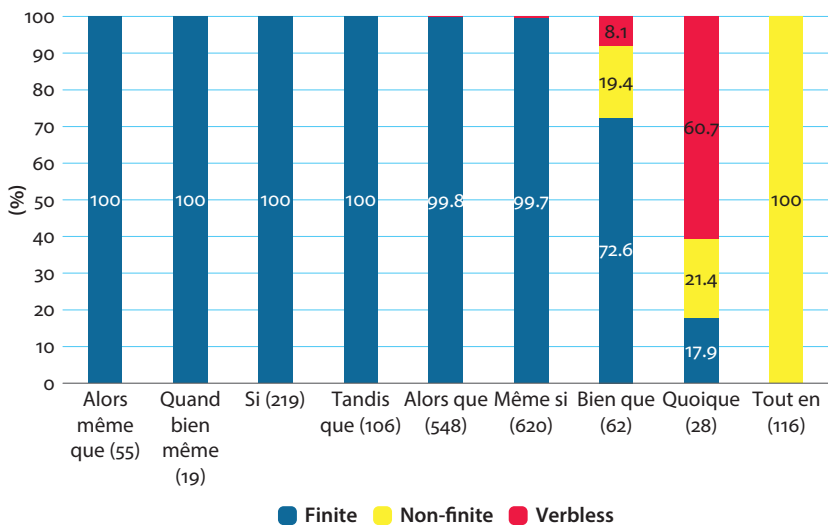
Figure 23. Breakdown of the clause types per subordinator in French (in percent)

Quoique and *bien que* are also among the few French subordinators that can be used in non-finite clauses, as illustrated in (131) and (132). Another subordinator, French *tout en*, can only be used in non-finite clauses, as exemplified in (133). This is unsurprising, as French *en* can only introduce present participles (Grevisse & Goosse 2013: 1405). It is notable that all the uses of *tout en* identified in the corpus combine temporal and contrastive meanings, which echoes the observation made on the non-finite uses of *while* in the previous paragraph.

(131)   **Bien que** disposant du temps nécessaire, Clinton n'est pas parvenu à transformer l'essai, à faire la paix en Palestine.   (Mult-Ed-FR – Le Figaro)

(132)   Jean Daniel, **quoique** régulièrement insulté par lui, a fait son éloge.
(Mult-Ed-FR – Le Nouvel Observateur)

(133)   Les usagers, **tout en** étant fiers de leurs services publics, aimeraient voir leurs préoccupations prises en compte, et n'être pas infantilisés dans des rapports défensifs avec des fonctionnaires parfois trop arrogants.
(Mult-Ed-FR – Le Figaro)

Finally, evidence of the diverging syntactic preferences of CMs also emerges from the analysis of the rank status of English CAs of contrast. As stands out from Figure 24, the propensity to occur in either hypotactic, embedded or minor clauses varies, sometimes markedly, across English conjunctive adjuncts. Whereas some markers (e.g. *meanwhile, even so*) never occur in any of these clauses, other CAs display some degree of variation with respect to the dominant pattern for English CAs, viz. main clauses.

Anyway, for example, seems to be particularly likely to occur in subordinate (i.e. both hypotactic and embedded) host segments, which together account for nearly 55% of its uses. Hypotactic and embedded uses of *anyway* are exemplified in (134) and (135), respectively, where the host clauses are underlined. Other markers that are commonly used in subordinate structures include *instead, still* or *on the other hand* – as in (136), (137) and (138), where *still* appears in an embedded clause, whereas the other two markers are used in hypotactic clauses. Figure 24 also suggests that minor clauses seem to be typical of only a small number of CAs. However, as the percentages of use of minor clauses reported in the graph correspond to very small numbers of occurrences in the corpus (e.g. the 5% of uses of *on the other* introducing minor clauses in fact correspond to a single occurrence in the corpus), this result is not discussed in detail here.

(134)   <u>Although the assault on Iraq would have been launched **anyway**</u>, the prospect of securing unencumbered access to an alternative oil source strengthened the pro-war argument.   (Mult-Ed-EN – The Guardian)

(135)   Quite often, the BBC has used licence-payers' money to outbid its independent rivals for films and sports events <u>that would have been shown free-to-air **anyway**</u>.   (Mult-Ed-EN – The Telegraph)

(136)   We have consistently called on the Government to seek alternatives to conflict with Iraq, and we will continue to do so, <u>urging **instead** that Britain maintains a policy of deterrence and containment</u>.
(Mult-Ed-EN – The Independent)

(137)   For those of us <u>who</u> queued three or four hours to get tickets to see Simon Russell Beale at the London's Donmar Theatre and <u>**still** failed to get any</u>, it's little consolation to know that he has gained the Evening Standard award for best actor and Sam Mendes a statuette for directing him.
(Mult-Ed-EN – The Independent)
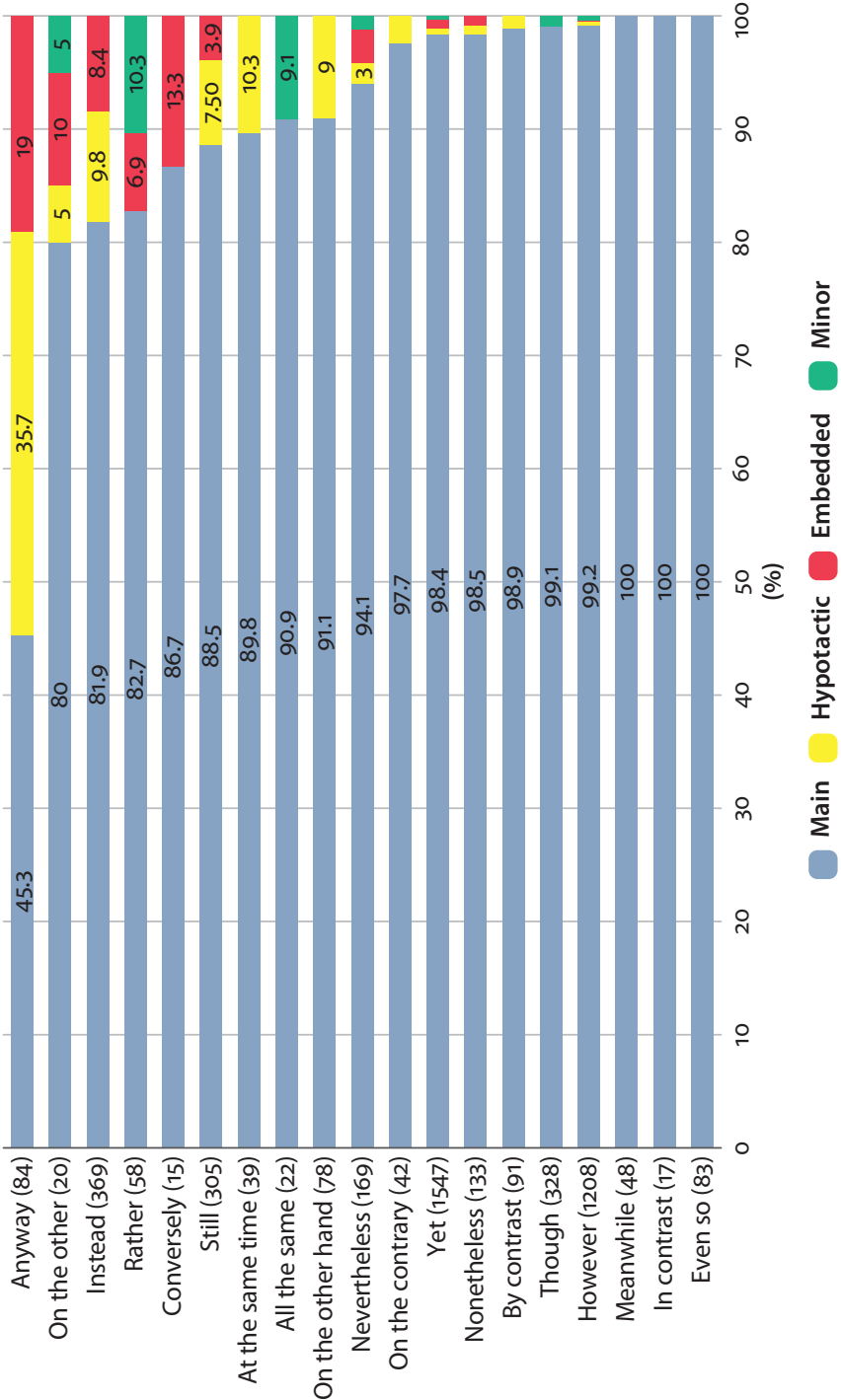
Figure 24. Breakdown of the rank statuses per conjunctive adjunct in English (in percent)

(138)   If several other states had voted "yes", europhiles would be able to trot out their old bus-missing metaphors. <u>If, **on the other hand**, another nation voted "no"</u>, Mr Blair would be off the hook.   (Mult-Ed-EN – The Telegraph)

In summary, the results presented in Figures 22 to 24 highlight that in both English and French, the different CMs do not all behave in the same way, but rather tend to differ in the kinds of host clauses which they prefer and/or allow for. All in all, the results presented in this section are thus in line with Hoey's (2005) theory of lexical priming (see above): the different English and French CMs of contrast appear to display their own specific syntactic features, which are not necessarily transferrable to the other members of their class.

Another way of looking at the corpus results is to start from the syntactic patterns themselves (instead of the lexical items) and determine whether a given pattern of use is evenly spread across CMs, or is rather attributable to one or two markers. Looking at the results from this perspective makes it possible to qualify the general tendencies underlined in the previous sections, by establishing whether they are representative of the language as a whole, or specific to one or two markers displaying eccentric usage patterns. It ought to be noted that the figures discussed here need to be treated with some caution. The fact that a large proportion of the CMs in a given category corresponds to a single lexical item is also partly a function of the frequency of occurrence of that marker in the corpus (e.g. as *pourtant* is nearly ten times as frequent as *néanmoins* in the corpus, it is also much more likely to display a higher frequency of verbless or minor clauses than *néanmoins*). In spite of this, approaching the data from this angle provides illuminating insights into the influence that lexis may have on the choice of a particular type of host segment. Looking at the cases of English CAs occurring in non-finite clauses, for instance, it appears that *instead* takes up no less than 64.2% of the 81 occurrences used in this pattern – as in Example (139). In addition, 75% of the 115 English CAs occurring in hypotactic clauses are accounted for by only three markers (out of nineteen), viz. *instead* (31.3%), *anyway* (26.1%) and *still* (17.4%), and nearly 80% of the 204 subordinators introducing non-finite clauses in English correspond to *while* – as in (140). Likewise, in French, nearly half of the 254 CMs used in hypotactic clauses correspond to *pourtant* (45.7%), whose hypotactic use is exemplified in (141). With respect to French subordinators, 87% of the 111 non-finite uses are instances of *tout en* – as in (142); while more than 85% of the 108 cases of minor clauses are introduced by *même si* (54.4%) or *alors que* (31.5%) – as in (143) and (144).

(139)   School kitchens have stopped cooking fresh food, becoming, **instead**, processing units for frozen junk.   (Mult-Ed-EN – The Observer)

(140)   Since replacing his elder brother as president in 2008, Raul Castro has
begun a series of economic reforms. **While** meant to "update socialism"
they have introduced elements of a market economy.

(Mult-Ed-EN – The Financial Times)

(141)   C'est une des affaires les plus révoltantes qu'on puisse trouver dans
l'actualité, qui en est **pourtant** prodigue.

(Mult-Ed-EN – Le Nouvel Observateur)

(142)   Quel bonheur, en tout cas, de retrouver la présence de François Furet,
notre ami, **tout en** regrettant, une fois encore, qu'une armée mandarinale
continue de sous-estimer l'intérêt et l'apport d'un aspect de son œuvre
qui a enrichi pendant des années les pages de notre journal.

(Mult-Ed-FR – Le Nouvel Observateur)

(143)   Ce jeune homme de 27 ans n'est pas le seul Français à s'être engagé dans le
djihad. **Même si** les autres, comme Hervé Loiseau, retrouvé mort sur les
contreforts de l'Hindu Kuch, ont souvent des origines arabes.

(Mult-Ed-FR – le Figaro)

(144)   [A]ussi surprenant que cela soit pour des démocraties occidentales où les
oubliés de la prospérité se réfugient dans l'abstention, deux tiers des ruraux
indiens, ceux-là mêmes qui survivent avec un dollar par jour et qui sont
bien souvent illettrés, ont exprimé leur mécontentement dans les urnes.
Alors que les nouvelles classes moyennes ne se sont bien souvent pas
dérangées pour aller voter.   (Mult-Ed-FR – Libération)

All these results suggest that the cross-linguistic differences uncovered in the pre-
vious sections may not be due (solely) to differences between the language systems
taken as wholes, but may also result from more specific lexical preferences of indi-
vidual CMs. A case in point concerns CAs used in hypotactic clauses: although in
Section 7.3.1, this pattern was shown to be markedly more frequent in French than
in English, a lexis-sensitive analysis of the results reveals that this difference largely
stems from the tendency of *pourtant* to occur in this type of structure.

Lexis-sensitive syntactic analyses such as the ones carried out here may be
highly relevant for foreign-language teaching and translator training. In peda-
gogical grammars and textbooks, conjunctive markers are frequently presented
as a fairly homogeneous set of discourse-structuring devices. While these books
typically make distinctions between the markers with respect to meaning (viz.
contrast, addition, cause, etc.) and grammatical category (viz. coordinators, sub-
ordinators, CAs), they hardly ever provide any indications on the actual usage
patterns – syntactic or otherwise – of these items. Yet, the corpus results obtained
here demonstrate that within the categories of subordinators and CAs, conjunc-
tive markers do not all behave in the same way. This implies that it is not enough

to simply specify the grammatical category to which each CM belongs, and expect learners or apprentice translators to simply transfer their knowledge of the general syntactic features of the category (e.g. subordinators may introduce either finite, non-finite and verbless clauses, see e.g. Quirk et al. 1985: 992) to the markers that constitute it. Instead, in order to ensure a maximally idiomatic, target-like use of conjunctive markers of contrast, teachers should sensitise learners to the fact that, depending on the specific CM that is chosen, they might have to make partly different choices regarding the syntactic features of the host clause. Thus, for example, it might be useful to make learners aware of the fact that *while* allows for the use of non-finite and verbless host clauses, whereas *even though* and *whereas* cannot be used with these structures, despite the fact that all three markers belong to the same grammatical category of subordinators. Admittedly, this is information that ought to be provided at fairly advanced levels of learning.

In addition, analysing the individual syntactic features of CMs may also help explain some translation tendencies, such as the lack of translation equivalence between CMs which are generally perceived as equivalent. In a previous study carried out with Sandrine Zufferey (see Dupont & Zufferey 2017), for example, we found that although *yet* and *pourtant* are usually considered to be cross-linguistically equivalent, they display a surprisingly low score of mutual correspondence. In the Europarl translation corpus of parliamentary debates, *yet* and *pourtant* are translated by each other in only 33% of the cases (ibid.: 282). In fact, this pair of CMs exhibits a rather strong translation bias, since *yet* is translated by *pourtant* nearly twice as frequently as the opposite (viz. 42% vs 24%). In cross-linguistic research, such a lack of equivalence is most frequently accounted for by reference to very fine-grained semantic differences between the markers, such as the degree of subjectivity of the related segments, their information status, or the range of concessive nuances that they may encode (see e.g. Zufferey & Cartoni 2012; Fretheim & Johansson 2002; see also Chapter 3 for an overview). In the case of *pourtant* and *yet*, however, syntax also seems to play a part in explaining the patterns of translational correspondence between the two markers. Comparing the syntactic features of these CMs, we observe that, whereas *yet* occurs in main clauses 98.4% of the time, French *pourtant* introduces main clauses in only 78.5% of the cases. Among the remaining 21.5% of the occurrences of *pourtant*, the dominant pattern is the hypotactic one (11.3%).

When *yet* is used in a main clause, it is perfectly possible to translate it by *pourtant*, which also very commonly appears in main clauses. This is indeed what happens in Example (145), extracted from the English-French component of the Europarl translation corpus. By contrast, in most cases when *pourtant* is used hypotactically, as in (146) and (147), *yet* can hardly be considered as an equivalent of *pourtant*, since *yet* is virtually restricted to syntactically independent clauses. Thus, unless the original examples undergo a complete syntactic restructuring,

translators will need to either select an equivalent which can in fact be used in a hypotactic clause – such as *nonetheless* in Example (146) – or simply delete the marker in the target text – as in (147). The same goes for embedded and phrasal-clausal uses of *pourtant*, which together account for nearly 10% of the occurrences of the marker. Therefore, the greater syntactic flexibility of French *pourtant* may partly explain why it is less easily rendered by the fairly 'rigid' *yet* than the opposite.

(145)   EN. **Yet** these 37 million people feel excluded and irrelevant to the European Union.

FR. **Pourtant** ces 37 millions de personnes se sentent exclues de l'Union européenne, insignifiantes à ses yeux.   (Europarl EN > FR)

(146)   FR. Elles imposent des réductions drastiques des coûts, toujours au détriment des salariés, <u>qui jouent **pourtant** un rôle essentiel dans le domaine de l'entretien et de la sécurité</u>.

EN. They demand drastic reductions in costs, always to the detriment of the employees, <u>who **nonetheless** play an essential part in the field of maintenance and safety</u>.   (Europarl FR > EN)

(147)   FR. La résolution commune aboutit à la condamnation en priorité du gouvernement de Serbie, <u>qui n'est **pourtant** en rien concerné par la question du retour des réfugiés</u>.

EN. The common resolution has led to a priority condemnation of the Serbian Government, <u>which is [X] not at all concerned by the question of returning refugees</u>.   (Europarl FR > EN)

In conclusion, the analysis of the syntactic patterning of contrastive CMs has shed light on two main significant factors of influence (in addition to the differences between the two language systems), viz. (i) discourse – since different syntactic uses of CMs have been shown to go hand in hand with a range of stylistic and rhetorical effects at discourse level; and (ii) lexis – since in both English and French, syntactic variation has been uncovered across CMs belonging to the same grammatical category. One question that arises from these results is the following: what factor, of lexis and discourse, has the greater influence on the syntactic features of the clauses hosting CMs of contrast? Given that some of the patterns giving rise to well-defined discourse effects (e.g. verbless uses of subordinators, or non-finite and hypotactic uses of conjunctive adjuncts) have also been shown to be mostly associated with specific conjunctive markers, it may be postulated that lexis is more influential than discourse in determining the syntactic features of the host clause. Alternatively, it may be hypothesised that some markers in particular have special potential to produce discourse effects. Further research is required to determine precisely what part is played by each of these factors. In any case,

however, the analyses presented in this chapter have provided convincing evidence that both lexis and discourse play a significant role in determining the syntactic patterns of English and French conjunctive markers of contrast.

7.4   Conjunctive adjuncts of contrast in English and French:
A cross-register comparison

So far, the cross-linguistic comparisons carried out in this chapter have been restricted to a single register, viz. newspaper editorials. However, a number of corpus studies in the monolingual field have demonstrated the highly register-sensitive nature of conjunctive markers. For example, Biber et al. (1999: 880–890), Conrad (1999) and Liu (2008) all reported significant variation in (i) overall frequencies of CMs; (ii) preferred semantic and syntactic types of CMs; and (iii) frequencies of individual CMs across the four English registers of academic writing, conversation, fiction and news. Smith and Frawley's (1983) comparison of English CMs in the fiction, news, religion and science written registers, and Altenberg's (1984; 1986) comparisons of causal and contrastive CM use in English speech and writing, yielded very similar results. Accordingly, and in line with contrastive linguists' growing concern for the impact of register variation on the differences between languages (see Chapter 3), Section 7.4 seeks to assess whether the cross-linguistic tendencies uncovered in Mult-Ed are stable across registers, or partly depend on the communicative situation investigated. For that purpose, the analyses carried out in Mult-Ed are complemented with similar analyses in the LOCRA comparable corpus of research articles. Little research has yet undertaken to compare the use of CMs in two or more languages across more than one register. Notable exceptions include the study by Kunz and Lapshinova-Koltunski (2015) on English and German, and the article by Taboada and Gómez-González (2012) on English and Spanish. Whereas in Kunz and Lapshinova-Koltunski's (2015) study, language emerges as the more influential factor in determining CM use, with cross-linguistic differences remaining relatively stable across registers, Taboada and Gómez-González (2012) show that differences in use of CMs are more marked between registers than between languages. In this section, I attempt to identify which of these two factors appears to be the more influential on the use of English and French CMs of contrast.

In view of the highly time-consuming character of the corpus analyses performed in the present chapter, the investigations presented in this section are more limited than those discussed so far. Firstly, with respect to CM frequencies of use, the register comparison is restricted to the sole category of conjunctive adjuncts (Section 7.4.1). Secondly, in order to assess the influence of register on the syntactic patterning of English and French CMs of contrast, I focused on the coordinators

but and *mais*, and more particularly on the proportion of sentence-initial coordinators used in each register (Section 7.4.2). This specific phenomenon was chosen because it has been demonstrated (i) to reflect the degree of formality of the texts in which it occurs, with sentence-initial uses being associated with speech-like or partly informal communication, and (ii) to contribute to creating well-defined effects at the discourse level, in line with the communicative goals of the texts in which it appears. As a result, this phenomenon is likely a good candidate to evaluate the impact of register on the syntax of CMs.

7.4.1 Frequency of conjunctive adjuncts of contrast in English and French: A comparison of newspaper editorials and academic writing

7.4.1.1 *Overall frequencies of English and French conjunctive markers in LOCRA and Mult-Ed*

In the second section of this chapter, the hypothesis that French tends to use more explicit conjunctive markers than English was tested in the Mult-Ed corpus of newspaper editorials. The corpus results led to a strong rejection of the initial hypothesis, as CMs of contrast were found to be significantly more frequent in English than in French, across all three syntactic categories of markers (viz. coordinators, subordinators and conjunctive adjuncts). Focusing on the category of conjunctive adjuncts, this section aims to determine whether the same kind of frequency difference is found in a different register (i.e. research articles), or whether the significant gap in frequency of CMs is (partly) specific to the editorial register. The relative frequency of conjunctive adjuncts in each subpart of the LOCRA corpus is provided in Table 18, where the frequencies of CAs in Mult-Ed are also reproduced. The table shows that, as was the case in Mult-Ed, conjunctive adjuncts of contrast occur significantly more frequently in the English than in the French subpart of LOCRA ($\chi^2 = 72.5$, $df = 1$, $p < 0.001$, $\varphi = 0.004$). However, the frequency difference in LOCRA is far less marked than the difference emerging from the Mult-Ed corpus. Whereas in Mult-Ed, English conjunctive adjuncts were 1.7 times as frequent as their French counterparts, in LOCRA this ratio drops to 1.2. Statistically speaking, the effect size associated with the chi-square value obtained for the Mult-Ed corpus ($\varphi = 0.01$) is larger than its LOCRA equivalent ($\varphi = 0.004$). In summary, what these results show is that the difference in frequency of CMs is stable across the two registers investigated here: in both editorials and scientific articles, English appears to use significantly more CMs than French – thus contradicting the claims commonly found in the English-French contrastive literature. In this, the results are in line with Kunz and Lapshinova-Koltunski's (2015), who identified language as the more determining factor with respect to CM use in English and German (as opposed to register). However, while register has no influence on the direction of the difference between English and French, it does

have an impact on the size of this difference, which is noticeably more marked in the editorial than in the academic register.

Table 18. Relative frequency of English and French conjunctive adjuncts of contrast in the Mult-Ed and the LOCRA corpora (per million words): cross-linguistic comparison

	ENGLISH	FRENCH
MULT-ED	2,320 (+16.6)	1,342 (−16.2)
LOCRA	2,335 (+6.0)	1,944 (−6.0)

Another illuminating contrast between English and French can be gleaned from the cross-register comparison of CM frequencies *within* each language. Table 19 reproduces the figures provided in Table 18, but with standardised residuals resulting from the chi-square comparison of frequencies of CMs across registers within each language (e.g. Mult-Ed-FR vs LOCRA-FR), instead of the comparison of CM frequencies across languages for each register (e.g. LOCRA-EN vs LOCRA-FR).

Table 19. Relative frequency of English and French conjunctive adjuncts of contrast in the Mult-Ed and the LOCRA corpora (per million words): cross-register comparison

	MULT-ED	LOCRA
ENGLISH	2,320	2,335
FRENCH	1,342 (+10.8)	1,944 (−10.6)

What Table 19 shows is that, whereas there is a very large, statistically significant difference in frequency of CMs between the French editorials and the French research articles ($\chi^2 = 229.0$, $df = 1$, $p < 0.001$, $\varphi = 0.007$), the two English registers display a nearly identical, and thus non-significantly different frequency of conjunctive adjuncts of contrast ($\chi^2 = 0.9$, $df = 1$, $p = 0.76$). In other words, what the comparison of academic writing and newspaper editorials in the two languages seems to suggest is that, when it comes to signalling relations of contrast explicitly by means of a CA, French seems to be more sensitive to register variation than English, which signals relations of contrast by means of explicit conjunctive adjuncts with a roughly equal frequency in the two registers. These results are in line with Crible (2018: 85), who reported that the frequency of discourse markers is more affected by register variation in French than in English spoken language based on the analysis of eight registers. In the light of Crible's result, we may postulate a generally greater register stability of English as compared to French regarding the use of metadiscourse markers. It is clear, however, that with respect to writing such a conclusion remains tentative at this stage: it only relies on the

comparison of two registers and therefore does not allow for any general claims on differences in register sensitivity between English and French.

The results obtained for English may appear somewhat surprising in view of the findings available in the previous literature. As explained above, earlier research comparing CM use across various English registers unanimously reported marked frequency differences depending on the communicative situation (see e.g. Biber et al. 1999: 880; Smith & Frawley 1983: 350; Conrad 1999: 7; Liu 2008: 499; Hempel & Degand 2008: 690). In particular, Biber et al. (1999: 880), Conrad (1999: 7) and Liu (2008: 500) all demonstrated that academic writing exhibits markedly higher frequencies of conjunctive adjuncts than journalistic prose, both overall, and with respect to the semantic category of contrast (with CMs of contrast being about twice as frequent in academic writing as in the editorials). According to Conrad (1999: 10), the low frequency of conjunctive adjuncts in news as compared to academic prose results from the fact that newspaper writing is less argumentative than academic writing:

> [News] discourse often follows a narrative organization. The point of the discourse may simply be the sequence of events, not the relationships between ideas, or the logical connections between events may be clear from the order of the sentences. […] The news excerpt reports the events without developing arguments, as academic prose typically does.

In a similar vein, according to Biber et al. (1999: 882), the frequency difference between the two registers stems from the fact that "reports of events are more common and arguments are rarer in news, and thus fewer linking signals are needed than in academic prose".

However, Biber et al. (1999), Conrad (1999) and Liu's (2008) news corpora are far from fully comparable to the editorial corpus used in the present study: editorials represent only a minor proportion of these researchers' data sets, along with more 'regular' news reports including world news, business news, arts news, sports news, etc. (see e.g. Biber et al. 1999: 31–32). The Mult-Ed corpus, by contrast, is exclusively made up of editorials, which are characterised by a strong argumentative component in addition to the informational goal that is central to most other news subgenres (see Chapter 5). In fact, in Biber's (1988: 149) corpus-based comparison of twenty-two spoken and written registers of English, editorials emerge as the second most persuasive text type. They are even shown to display a markedly higher frequency of linguistic features typical of persuasive language than academic prose. Moreover, according to Biber (ibid.: 150), one of the most typical manifestations of argumentation in editorials is the fact that in those texts, "[s]everal perspectives are considered, with arguments for and against them, but the overall discourse builds towards a final conclusion and attempts to convince the reader that this conclusion is superior to any other". Strikingly, Biber's

description of this textual strategy is more or less a paraphrase of the definition of concession provided in Chapter 2. Since they are in perfect accordance with the kind of argumentative moves typical of editorials, it is not surprising that CMs of contrast (and of concession, in particular) should be frequent in this register. Finally, according to Hempel and Degand (2008: 683), one should expect more organisational metadiscourse in newspaper articles that are "politically or ideologically tinted, as the author wishes to convince his/her readers of the relevance of his/her point of view". In view of all this, the absence of a frequency difference between LOCRA-EN and Mult-Ed-EN, along with the discrepancy between the results obtained here and previous cross-register comparisons, most probably result from the argumentative character of the texts included in Mult-Ed.

If this is indeed the case, however, it now becomes difficult to interpret the significantly higher frequency of French CMs in academic prose than in newspaper editorials – a result which actually matches the difference found by Biber et al. (1999), Conrad (1999) or Liu (2008) between academic prose and the 'general' register of newspaper writing. Given that the communicative features of the two registers studied here are supposedly comparable in the two languages (see Chapter 5), it is astonishing to find such a marked difference between English and French as regards the impact of register on CM use. Such results may reflect differences in the conventions of the academic and/or editorial registers in English and French. For example, given that conjunctive markers have (partly) been associated with argumentative language, a less persuasive, more factual tone in French editorials as compared to their English counterparts might explain why editorials do not reach the CM frequency of academic writing in French, whereas they do in English.

Another possible explanation for the frequency gap between the two French registers may have to do with the difference in degree of formality of the two text types. In Section 7.2, a number of alternative means used by French writers to express contrast were put forward as a way to explain the very striking gap in frequency of CMs between English and French. Those included thematising syntactic structures (e.g. *Pierre, il…; c'est Pierre qui…*), emphatic pronouns (e.g. *moi, lui, eux,* etc.) and a number of disjuncts combining a stance meaning with a contrastive one (e.g. *malheureusement, bien sûr, à vrai dire, certes,* etc.). Some of these alternative strategies, however, are more typical of speech or informal language. Rivelin-Constantin (1992: 159), for example, explains that thematisation is much more frequent in oral than in written French (see also Trévise 1986: 196). As already discussed on several occasions in this chapter, there is evidence that in recent decades, newspaper language has been evolving towards a more informal, speech-like style of expression. Academic writing, by contrast, has not undergone such a change, but has tended to remain a resolutely formal register (see Hundt & Mair 1999; Hyland & Jiang 2017; see below for a more detailed discussion). In view of this, it may be hypothesised that, whereas some relations of contrast

are expressed through thematisation or emphatic pronouns in French editorials, the formal nature of research articles prevents writers from using such resources, leading them to use conjunctive markers instead. It is clear that, at this stage, these suggested explanations remain largely speculative. Further research is required in order to pinpoint exactly what factors may explain the striking frequency difference between French editorials and research articles. In addition, it is also conceivable that the frequency difference between the two French registers may be traceable to one or two markers, which would be highly typical of the type of argumentation found in academic discourse, but much less frequent in editorial writing. This hypothesis is checked in the following section.

7.4.1.2 *Lexical breakdown of the corpus results*

Examining the overall frequencies of CMs is only part of a cross-register comparison of discourse organisation. Registers that express a given discourse relation with roughly the same frequency (as is the case for English editorials and research articles) may very well do so with different means. This has been demonstrated in previous corpus-based research, which has underlined marked differences in the frequencies of use of individual markers across registers. The relative frequencies of each individual CM per register are provided in Table 20 for English, and Table 21 for French. The numbers provided between brackets next to each CM type in LOCRA indicate the rank of this type in the Mult-Ed corpus, in order to get a clear picture of possible differences in importance of the CMs across registers. CM types that only occur in one of the two registers are underlined in the tables. Again, CMs whose frequency did not reach the threshold of five occurrences per million words are excluded from the analysis.

Table 20 shows that, whereas the English subsets of Mult-Ed and LOCRA did not exhibit any significant difference in overall frequency of CMs, the two registers display some striking differences in the items which they tend to prefer to express contrast. For example, although the top three CMs are the same in the two corpora (viz. *however, yet* and *instead*), the most frequent marker differs according to the register: while *yet* is the preferred conjunctive adjunct in Mult-Ed, *however* occupies the first position in LOCRA. In addition, the gap between the first and the second markers is much wider in LOCRA (where *however* accounts for exactly 50% of the total number of occurrences) than in Mult-Ed (where *yet* represents 33% of the total number of CAs). Thus, *however* seems to be particularly well-suited to the discourse purposes of the academic register: it constitutes the preferred choice in half of the cases in which writers wish to express a contrast between two clauses or sentences by means of a CA. Even though it ranges among the most frequent CAs in Mult-Ed (where it accounts for 26% of the occurrences), it is nevertheless nearly twice as infrequent in that corpus as it is in LOCRA.

Table 20. Relative frequency of individual English conjunctive adjuncts of contrast in the Mult-Ed and LOCRA corpora (per million words)

	Mult-Ed		LOCRA	
	Conjunctive Marker	**Relative Frequency (pmw)**	**Conjunctive Marker**	**Relative Frequency (pmw)**
1	Yet	771	However (2)	1,169
2	However	602	Yet (1)	244
3	Instead	184	Instead (3)	181
4	Though	163	Rather (12)	126
5	Still	152	In contrast (18)	102
6	Nevertheless	84	On the other hand (11)	95
7	Nonetheless/none the less	66	Nevertheless (6)	93
8	By contrast	45	Still (5)	80
9	Anyway	42	Nonetheless/none the less (7)	68
10	Even so	41	By contrast (8)	58
11	On the other hand	39	Conversely (19)	33
12	Rather	29	Though (4)	26
13	Meanwhile	24	At the same time (15)	24
14	On the contrary	21	On the contrary (14)	21
15	At the same time	19	Anyway (9)	7
16	All the same	11	Meanwhile (13)	6
17	On the other	10		
18	In contrast	8		
19	Conversely	7		

In fact, what Table 20 shows is that some CMs occur with similar frequencies in the two corpora, whereas others appear to be more typical of – or even exclusive to – one of the two registers. This distinction corresponds to the one between 'style-neutral', 'style-biased' and 'style-restricted' markers made by Altenberg (1986: 17) in his comparison of CMs of contrast in spoken and written English. In the English data, seemingly style-neutral CMs (i.e. whose frequencies do not differ significantly between the two registers, as revealed by chi-square tests of independence) include: *instead, nevertheless, nonetheless, on the contrary, by contrast* and *at the same time.* In contrast, a range of other CMs appear to be biased in favour of one of the two registers, therefore being more style-sensitive. In Mult-Ed, these CMs are *though, still* and *anyway* – all of which are usually associated with a more

informal, speech-like style of expression (see e.g. Altenberg 1986: 17; Biber et al. 1999: 887). The higher frequency of these CMs in the editorial corpus is therefore likely attributable to a difference in degree of formality between the two registers. Research has shown that, whereas newspaper language has progressively been integrating a number of colloquial linguistic features over the past decades, academic prose did not undergo such an evolution along the formality continuum. For example, based on a diachronic comparison of journalistic and academic texts along a range of typically spoken and/or informal linguistic phenomena (e.g. sentence-initial coordinators, contractions, voice, progressives, abstract nouns, etc.), Hundt and Mair (1999: 236) conclude that academic prose can be viewed as a rather conservative, or even 'uptight' genre as compared to the more 'agile' newspaper register, in the sense that it is far less permeable to speech-like linguistic features. Likewise, in a study seeking to assess the degree to which academic prose has been integrating features of informal language over the past fifty years, Hyland and Jiang (2017) investigate a range of linguistic phenomena typically associated with informality across four scientific disciplines (viz. applied linguistics, sociology, biology and electrical engineering). They report that, whereas the hard sciences have been evolving in the direction of a more informal style of expression, the results obtained for the Humanities – i.e. articles which are comparable to those included in the LOCRA corpus – "indicate a reduction in the use of these informal features" (ibid.: 48). Academic writing thus remains a predominantly formal genre, which normally precludes the use of such informal markers as *anyway* or adverbial *though*.

Finally, the corpus results also shed light on CMs that are exclusive to the editorial register, and never occur in the academic corpus, viz. *even so, all the same* and *on the other*. Note that these three 'style-restricted' items (to quote Altenberg) nevertheless represent a very small proportion (c. 2.5%) of the total number of CMs used in the editorial corpus. Academic writing does not appear to have any such style-restricted markers, since all the sixteen CMs extracted from LOCRA also occur in Mult-Ed. However, the LOCRA corpus displays some style-biased CMs, which are markedly more frequent than in the editorials, viz. *rather* (the fourth most frequent CM in LOCRA, which only occupies the twelfth position in Mult-Ed), *in contrast* (ranked fifth in LOCRA, and eighteenth in Mult-Ed) or *on the other hand* (sixth in LOCRA, and eleventh in Mult-Ed).

Table 21 provides the relative frequency of each French CM in the two registers. The first observation to be made on this table is that, in contradiction with the hypothesis formulated at the end of the previous section, there does not seem to be one or two CMs which account for the significant gap in overall frequencies of CMs between the editorials and the research articles. Rather, CMs of contrast seem to be generally more frequent in LOCRA than in Mult-Ed. For example,

Table 21. Relative frequency of individual French conjunctive adjuncts of contrast in the Mult-Ed and LOCRA corpora (per million words)

	MULT-ED		LOCRA	
	CONJUNCTIVE MARKER	RELATIVE FREQUENCY (PMW)	CONJUNCTIVE MARKER	RELATIVE FREQUENCY (PMW)
1	Pourtant	491	Cependant (3)	416
2	Or	170	Toutefois (6)	273
3	Cependant	135	Pourtant (1)	242
4	Au contraire	120	Néanmoins (8)	190
5	En revanche	110	En revanche (5)	187
6	Toutefois	65	Au contraire (4)	141
7	Plutôt	62	Or (2)	130
8	Néanmoins	52	Plutôt (7)	121
9	Tout de même	42	À l'inverse (14)	57
10	De l'autre	22	Par contre	52
11	Malgré tout	21	Inversement	27
12	De toute façon	13	Tout de même (9)	23
13	Bien au contraire	11	D'autre part	17
14	À l'inverse	10	Malgré tout (11)	14
15	Quand même	8	A contrario	12
16	En même temps	5	À l'opposé	11
17	De toute manière	5	De l'autre (10)	8
18			Bien au contraire (13)	8
19			En même temps (16)	7
20			D'un autre côté	6

apart from the first most frequent CM – which displays a lower frequency in LOCRA than in Mult-Ed – the top eight markers in LOCRA all occur markedly more frequently than the top eight CMs in Mult-Ed, viz. the second most frequent CM in LOCRA (273 pmw) is more frequent than the second most frequent CM in Mult-Ed (170 pmw), the third most frequent marker in LOCRA (242 pmw) is more frequent than the third most frequent CM in Mult-Ed (135 pmw), and so on. In other words, rather than being attributable to the highly frequent use of one or two very 'typically-academic' CMs of contrast, the difference between the two registers genuinely seems to correspond to a generally greater propensity of French academic writing to signal relations of contrast by means of explicit conjunctive adjuncts.

Table 21 also shows that in some respects, and as was the case for the overall frequencies, French seems to display more register variation than English. For example, whereas the top three English markers were identical in the two registers, in French only two of the three most frequent markers are shared by the registers. While *or* ranks second in Mult-Ed, it only appears in seventh position in LOCRA. Conversely, *toutefois* is the second most common CM in LOCRA, whereas it only comes sixth in Mult-Ed. In addition, the ranks of *pourtant* and *cependant* are reversed in the two registers – with *pourtant* being more typical of the editorials, and *cependant* being the preferred item in LOCRA. In this regard, the situation reflects the results obtained for English, where the ranks of *yet* and *however* – usually presented as equivalents to *pourtant* and *cependant*, respectively – were also inverted. It is therefore reasonable to assume that there is something about the meaning of *pourtant/yet* that is particularly well-suited for the type of argumentation found in editorials, whereas *cependant/however* is especially appropriate for the kind of developments typical of academic prose. In a future study, those markers would deserve a fine-grained semantic analysis to better grasp the origin of their register-sensitivity. A second element that suggests a greater degree of register sensitivity of French CM usage is the fact that French displays a larger number of style-restricted markers than English, i.e. markers which occur in only one of the two registers. Whereas in English, only three markers were restricted to the editorial register, French not only displays three editorial-specific markers (viz. *de toute façon, quand même* and *de toute manière*), but also six CMs occurring only in the academic subcorpus (viz. *par contre, inversement, d'autre part, a contrario, à l'opposé, d'un autre côté*). Again, these markers only represent a small proportion of the total number of CMs in each corpus (viz. 2% in Mult-Ed; 6% in LOCRA).

On the other hand, with respect to the distinction between style-biased and style-neutral CMs, the picture appears somewhat less clear than in English (where *though, still* and *anyway* were unmistakeably more typical of the editorials, and *in contrast, on the other hand* and *rather* were very clearly associated with the research articles). The waters are partly muddied by the fact that academic writing demands more CMs of contrast than the editorials overall, which blurs the boundary between relative frequencies and ranks. For example, *plutôt* is twice as frequent in LOCRA as in Mult-Ed, but occupies a higher rank in the editorial subcorpus. A similar situation is observed with *en revanche*. Consequently, all that can be said at this stage is that French appears to display few truly stylistically-neutral CMs: apart from *au contraire*, whose frequency does not differ significantly between the two text types, few of the frequent markers occur with

strikingly similar frequencies in the two subcorpora.[14] Among the most strikingly stylistically-biased CMs are *toutefois* and *néanmoins*, which are both much more typical of academic texts.

In summary, the lexical breakdown of CM frequencies per register has shown that in both English and French, the choice of CM types is largely dependent on the communicative situation in which they are used. In both languages, the corpus results shed light on (i) CMs which only occurred in one of the two registers investigated; (ii) CMs that were more strongly associated with one register (e.g. *yet, though* or *pourtant* in editorials; *toutefois* and *rather* in the research articles). In other words, as noted by Greenbaum (1969: 80), many conjunctive adjuncts appear to act as 'style markers'. On the other hand, the corpus analysis also uncovered CMs whose frequencies did not appear to be affected by register, especially in English (e.g. *instead, on the contrary, au contraire*).

7.4.2 Syntactic patterns of English and French conjunctive markers of contrast in editorials and academic writing: A focus on sentence-initial coordinators

This section aims to assess whether, in addition to influencing the frequencies and types of CMs of contrast used, register also has an effect on the syntactic patterning of these markers, and/or on the differences between English and French in that respect. For this purpose, the section focuses on the syntax of the coordinators *but* and *mais*, and more particularly on the intersentential uses of these items. In order to verify whether the (surprisingly) high ratio of intersentential coordinators uncovered in Section 7.3 was related to the features of the editorial register, I carried out a small-scale analysis of *but* and *mais* in LOCRA. I extracted random samples of 300 occurrences of *but* and *mais* from the academic corpus and disambiguated the data to only keep cohesive uses of coordinators occurring in main clauses. This left me with 143 relevant instances of *but*, and 135 occurrences of *mais*. The percentages of inter- and intrasentential uses of *but* and *mais* in LOCRA are provided in Table 23, whereas Table 22 reproduces the results obtained in the editorial corpus.

---

14. While there is no statistically significant frequency difference between the registers for the markers *malgré tout, bien au contraire* and *en même temps*, the frequencies of those items in the two subcorpora is felt to be too small to allow for any definite conclusions on their degree of register sensitivity.

Table 22. Inter- and intrasentential uses of English and French coordinators of contrast in Mult-Ed (in percent)

	ENGLISH	FRENCH
INTERSENTENTIAL USES	61% (−1.83)	69.5% (+1.99)
INTRASENTENTIAL USES	39% (−2.23)	30.5% (+2.29)
TOTAL (RAW)	1541 (100%)	1309 (100%)

Table 23. Inter- and intrasentential uses of English and French coordinators of contrast in LOCRA (in percent)

	ENGLISH	FRENCH
INTERSENTENTIAL USES	22.4% (−2.23)	44.4% (+2.29)
INTRASENTENTIAL USES	77.6% (+1.57)	55.6% (−1.61)
TOTAL (RAW)	143 (100%)	135 (100%)

As appears clearly from these tables, sentence-initial coordinators of contrast are markedly less frequent in academic writing than in newspaper editorials, whether we look at the English or the French data. English intersentential *but* is nearly three times as frequent in the editorials (61%) as in the research articles (22.4%). In French, on the other hand, the cross-register difference is slightly less striking, with editorials displaying just over one and a half times more sentence-initial *mais* (69.5%) than academic prose (44.4%). In other words, while sentence-initial position was the preferred option in Mult-Ed, in the research articles, the unmarked pattern appears to be the intrasentential one. This means that in both languages, academic writing tends to be more compliant than newspaper language with the longstanding prescriptive rules which condemn the use of coordinators at the beginning of a sentence (see Section 7.3.1.2 for a discussion). The results are in line with Hundt and Mair (ibid.: 227–228), who show that the growing frequency of intersentential coordinators in English newspaper writing over the last third of the twentieth century is not observed in a comparable diachronic corpus of academic prose. Instead, academic writing displays a fairly stable, and comparatively much lower frequency of intersentential coordinators.

The cross-linguistic comparison reveals that, as was the case in the editorials, the ratio of inter- and intrasentential coordinators differs significantly between languages (χ^2 = 15.3, *df* = 1, p < 0.001, φ = 0.23), with French displaying a higher proportion of sentence-initial coordinators than English. In fact, the difference between the two languages is more pronounced in the academic (22.4% vs 44.4%) than in the editorial (61% vs 69.5%) register. This is visible from the standardised

residuals emerging from each chi-square comparison. Whereas in Mult-Ed, the standardised residuals only indicated a slightly significant overuse of intersentential coordinators in French (i.e. just over the 1.96 significance threshold), the values obtained for the LOCRA corpus indicate both that (i) English coordinators are significantly *less* likely to occur sentence-initially than expected, and (ii) French coordinators are significantly *more* likely to occur sentence-initially than expected. Thus, as was the case for the overall frequencies of conjunctive adjuncts, register influences not the actual difference between English and French (as French displays more intersentential coordinators than English in both registers), but the extent of this cross-linguistic difference, which is more substantial in one register than the other. It is also noteworthy that, whereas CM frequencies were shown to be more affected by register in French than in English, with respect to sentence-initial coordinators, it is English that seems to be more register-sensitive than French (since English displays a wider cross-register difference in frequency of intersentential coordinators than French). Admittedly, the samples analysed for the academic register are fairly small, and only represent a small portion of the patterns analysed in Section 7.3. Although it would be worth investigating other syntactic features in the academic register (e.g. the ratio of verbless, minor or hypotactic clauses), the pilot study on *but* and *mais* already provides evidence of the impact of register on the syntactic patterning of English and French CMs of contrast.

In Section 7.3, it was demonstrated that one of the reasons for using sentence-initial coordinators of contrast is argumentation: sentence-initial coordinators were shown to have the potential of bringing emphasis on both the relation of contrast, and the discourse segment introduced by the coordinator. As argumentation is an important component of the academic register (cf. above; see also Chapter 5), sentence-initial coordinators could function as a useful resource to strengthen some aspects of the writers' argumentative developments. However, the corpus results show that researchers seem to make little use of the argumentative potential afforded by sentence-initial coordinators, especially in English. In this case, the lower frequency of intersentential coordinators in academic writing as compared to editorials is most probably traceable to a difference in degree of formality of the two registers. As explained in Section 7.3.2, sentence-initial coordinators have typically been associated with informal or spoken forms of communication. Accordingly, studies which have noted a significant rise in frequency of such patterns in newspaper writing have generally interpreted this change as a manifestation of the gradual colloquialisation of newspaper language (e.g. Cotter 2003; Hundt & Mair 1999). Unlike newspaper writing, however, academic prose has been much more impermeable to such stylistic innovations, showing a tendency to remain situated towards the more formal end of the formality continuum – at least in the Humanities (see above for a more detailed discussion).

In summary, this section has provided convincing evidence of the influence of register on the frequency and usage patterns of English and French conjunctive markers of contrast, and on the differences between the two languages in this respect. Whereas register did not appear to have an impact on the *direction* of the differences between English and French (e.g. CMs of contrast were found to be more frequent in English than in French in both registers; sentence-initial coordinators were more typical of French than English in both LOCRA and Mult-Ed), it did affect the *size* of these cross-linguistic differences, which were more or less marked depending on the communicative situation. Once again, the results discussed in the section may be of great pedagogical relevance. A number of studies in the field of learner corpus research have reported cases of stylistic misuse of CMs in learner writing. In their study of CM usage in the essay writing of Cantonese learners of English, for instance, Field and Yip (1992: 26) show that "some of the devices used by L2 writers […] can give confusing signals of register to native English speakers", either by being too formal or too colloquial in view of the context in which they are used. Likewise, Granger and Tyson (1996: 23) explain that "learners […] seem to be unaware of [the] stylistic restrictions" associated with conjunctive adjuncts – as attested by an overuse of the informal markers *so* and *anyway* in the written productions of French learners of English. Altenberg and Tapper (1998: 88) also uncover "a general stylistic uncertainty about the use of connectors" in the argumentative writing of Swedish learners of English. These researchers have attributed such infelicities to a lack of pedagogical attention paid to the stylistic properties of conjunctive markers. Conrad (2004), for instance, highlights the paucity of information on register variation in foreign-language teaching materials. Based on a review of four recent ESL textbooks, she shows that, although conjunctive adjuncts are almost always covered in these manuals, they are typically unaccompanied by any comments on their preferred contexts of use. For example, she explains that, despite the fact that *though* is the most frequent conjunctive adjunct of contrast in her native-English corpus of conversation, only one in four textbooks mentions this marker. In addition, she shows that the single book which makes an explicit mention of the stylistic features of CMs advises learners to use *however* and *on the other hand* to express concession in speech. Yet, corpus results reveal these two markers are infrequent in conversation, and more typical of academic writing (ibid.: 73).

The results presented in this section may constitute a good starting point for teachers and pedagogues to sensitise learners to the variation not only in overall and individual frequencies, but also in syntactic patterning of conjunctive markers across communicative situations. Fortunately for both French-speaking learners of English, and English-speaking learners of French, some patterns of variation are somewhat similar across languages. For example, the CM pairs *yet/pourtant*

and *however/cependant* appear to share some stylistic properties, with the former pair being more typical of newspaper editorials, and the latter being especially frequent in academic writing. Likewise, sentence-initial uses of both *but* and *mais* are more frequent in the editorials than in the research articles – although French seems to be more tolerant of such uses in both registers. Other aspects of variation, by contrast, appear to be less stable across languages. The most striking example of this phenomenon is perhaps the amount of cross-register variation in overall frequencies of CMs uncovered in each language – where no significant difference was found between the two English registers, whereas the French editorials and research articles displayed a highly significant difference in frequency of CMs of contrast. Consequently, for example, English learners of French may need to be made aware that, depending on the type of writing that they have to produce, they might need to be more or less parsimonious in their use of explicit CMs of contrast. Conversely, it may be useful to inform French-speaking learners of English that, although informal CMs such as *though, anyway* or *still* are acceptable in *some* written registers (such as editorials), they are inappropriate in other forms of writing – including academic prose.

7.5   Conclusion

This chapter has investigated and compared the frequency and syntactic patterns of use of French and English conjunctive markers of contrast. Firstly, with respect to CM frequencies, the corpus approach adopted here has allowed me to test and, for the most part, contradict a number of longstanding, introspection-based statements made in the English-French contrastive literature. For example, the hypothesis that French tends to be more explicitly cohesive than English was strongly rejected by the corpus evidence. In fact, English was found to display a markedly higher frequency of explicit CMs of contrast than French – with a ratio of about 1.5 English CM of contrast per French marker. Likewise, the alleged preferences of English for coordinators, and French for subordinators and conjunctive adjuncts were, for the most part, unsupported by the corpus results. In terms of frequencies, English was found to use more CMs than French across all three grammatical categories. The analysis of proportions, on the other hand, only confirmed the preference of French for subordinators, whereas the proportion of conjunctive adjuncts was larger in English, and that of coordinators did not differ significantly between languages. It is clear that the corpus analyses carried out in this chapter only pertain to one single logico-semantic relation, viz. contrast. While these results do not allow for a categorical rejection of the claims available in the contrastive literature, they nonetheless clearly highlight the absolute necessity to refine these statements,

most of all by nuancing them according to the type of relationship which unites the discourse segments.

Alongside the comparison of CM frequencies, this chapter also investigated the syntactic features of the clauses typically hosting English and French CMs of contrast. Such an approach is uncommon in current research on conjunctive markers, to the point that it may have been felt to be irrelevant within the framework of the present study. Yet, the corpus results demonstrated the value of such a fine-grained description of CMs, which goes beyond the broad categorisation of the markers into coordinators, subordinators and conjunctive adjuncts. Despite the general dominance of CMs used in finite and main clauses (or hypotactic clauses, for subordinators) in both languages, the corpus results also pointed to a number of significant differences in syntactic patterning of English and French markers of contrast. Importantly, the various syntactic uses of conjunctive markers of contrast were shown to reflect a range of stylistic effects at the discourse level, thus being situated at the syntax-discourse interface. On the one hand, some patterns (including sentence-initial coordinators and CMs used in minor clauses) were shown to function as devices of textual fragmentation, which had the effect of emphasising both the discourse relation of contrast signalled by the CM, and the segment introduced by it. While such patterns were found in both languages, they were significantly more frequent in French. Thus, the differences in syntactic patterning of contrastive CMs were found to be revealing of diverging strategies of textual development in the two languages. Other syntactic patterns, by contrast, were shown to be associated with a certain degree of compression in the expression of conjunctive relations of contrast – in line with the space-saving imperatives of journalistic writing. Such uses were found in both English and French, although they had different manifestations in the two languages (i.e. syntactic compression was ensured by verbless uses of subordinators in English, and non-finite and/or hypotactic uses of coordinators and conjunctive adjuncts in French). Interestingly, the two main types of effects achieved by the syntax of CMs of contrast were also shown to pull in opposite directions, embodying a tension between (i) an evolution of newspaper language towards more informal, speech-like and reader-friendly styles of expression (a tendency which was in fact visible in a large number of the corpus results discussed in this chapter), and (ii) the use of more literate, less explicit writing techniques involving the compression of information within a minimum of words. In addition to the interaction between syntax and discourse, the corpus analysis also uncovered a close relationship between the syntactic patterns of use of CMs and lexical factors. It was established that in both English and French, the various CMs within each grammatical category do not all display the same syntactic features, but that some CMs exhibit a certain degree of idiosyncrasy. Accordingly, some of the cross-linguistic differences in syntactic patterning

were shown to be largely associated with small sets of CMs displaying distinctive syntactic features. In sum, CM use was thus demonstrated to lie at the interface between syntax, discourse and lexis.

A final factor of influence investigated in this chapter was register. In line with the recent developments of corpus-based contrastive linguistics, some of the analyses carried out in Mult-Ed were replicated in the LOCRA corpus of research articles, so as to assess the degree of generalisability of the differences uncovered between the languages. As was the case in the editorials, (i) conjunctive adjuncts of contrast were more frequent in English than in French; and (ii) sentence-initial coordinators were more common in French than in English. However, the size of these cross-linguistic differences was shown to vary in function of register. In addition, the two languages appeared to differ in the extent of register variation that they displayed. For example, whereas English showed no significant difference in frequency of conjunctive adjuncts in Mult-Ed and LOCRA, in French these CMs were significantly more frequent in the research articles than in the editorials. Finally, the register analysis also shed light on a number of interesting cross-register differences in CM use *within* each language system (see e.g. the distinction between style-neutral, style-restricted and style-biased CMs).

At a more general level, one of the important contributions of this chapter has been to show that analyses at the macro level do not prevent the study of more micro phenomena. Whereas the investigations carried out in this chapter started from a general category (viz. CMs of contrast), studied in a very large data set (viz. more than 4 million words in total, corresponding to several thousands of CM tokens), the discussion also tackled linguistic phenomena at very fine-grained levels of analysis (e.g. qualitative comments on the discourse effects produced by various syntactic uses of CMs of contrast; fine-grained differences in the individual syntactic patterning of English and French CMs of contrast). In fact, not only does the adoption of a macro perspective allow for micro-level analyses, but it also makes it possible to put these micro phenomena in perspective. For example, it is only by starting from the general class of CMs that I was able to identify differences in syntactic preferences between the markers analysed (e.g. the subordinators *albeit* and *whereas*, although belonging to the same grammatical category, were shown not to allow for the same types of host clauses). This is not something that could have emerged from a case study focusing on one or two markers only.

Placement patterns of English and French conjunctive adjuncts of contrast

8.1 Introduction

In Systemic Functional and other linguistic theories, conjunctive markers have often been said to be closely associated with the thematic – or initial – part of the message. On the one hand, coordinators and subordinators have been shown to be "inherently thematic" (Halliday & Matthiessen 2004: 83), as they may *only* occur at the beginning of a clause or sentence. Conjunctive adjuncts, on the other hand, have been described as "characteristically thematic" or as "natural Themes" (ibid.): due to their inherent linking function, they are generally placed at the beginning of the second segment involved in the discourse relation, so that its connection with preceding discourse can appear as clearly as possible to the reader or listener (see Chapter 4, Section 4.2.2 for more details on the close relationship between conjunctive markers and Theme). Yet, as opposed to coordinators and subordinators, conjunctive adjuncts are not *fixed* at clause boundary. Instead, in many languages (including English and French), conjunctive adjuncts such as *however* or *cependant* are syntactically mobile elements (see e.g. Biber et al. 1999: 890–892; Rubattel 1982: 50). This means that, in addition to clause- or sentence-initial position – as in Example (1) – they may also occur medially – as in (2) – or finally – as in (3).[1]

(1) **However**, my father didn't worry.

(2) My father, **however,** didn't worry.

(3) My father didn't worry, **however**.

Several factors have been shown to influence the placement patterns of conjunctive adjuncts in the literature. Firstly, as explained in Chapter 4, a number of contrastive studies have demonstrated that languages differ regarding the positions which they tend to prefer for conjunctive adjuncts, even when they have a similar set of syntactic options at their disposal. Altenberg (1998, 2006), for instance, found

---

1. The examples are adapted from Altenberg (2006: 12).

that while a majority of English conjunctive adjuncts occur in initial position, Swedish CAs are most commonly used sentence-medially. Kunz and Lapshinova-Koltunski (2014) demonstrated that English CAs are used initially more than twice as frequently as their German counterparts in a genre-varied comparable corpus of written texts. Likewise, Balažic Bulc and Gorjanc (2015: 64–66) show that Croatian conjunctive adjuncts occur in initial position significantly more frequently than their Slovene equivalents (viz. 96% vs 57%) in a comparable corpus of academic articles. Thus, depending on the language system investigated, the strength of the relationship between conjunctive adjuncts and thematic position is likely to vary, sometimes markedly.

Another factor that has been demonstrated to influence CA placement is lexis. A few studies have drawn attention to the fact that different CAs within a given language system do not necessarily display the same positional preferences. In his early study of adverbial positions in English, for example, Jacobson (1964) provides a corpus-based word-order dictionary of English adverbs, whose "main purpose is to show how individual adverbs are actually placed by present-day writers" (ibid.: 365). Based on the analysis of a collection of 66 books of both fictional and non-fictional written prose, he provides the proportion of use of each adverb in initial, medial and final positions. A comparison of the entries provided for various conjunctive adjuncts clearly shows that different CAs display partly divergent positional preferences: for example, while 76% of the occurrences of *nevertheless* are used sentence-initially, *therefore* is most often found in medial position (62%), with only a third of its occurrences used initially. Although enlightening, Jacobson's account is fairly limited empirically (e.g. only c. 80 occurrences of *therefore*; c. 50 occurrences of *nevertheless* are investigated). On a larger scale, and in a learner corpus framework, Paquot (2010: 178) compares the proportion of sentence-initial conjunctive markers in the written productions of learner and expert users of English. Her results show extensive variation in proportion of initial uses per CA in the expert corpus. For example, whereas *furthermore* is found initially in 81.1% of the cases, only 5.3% of the occurrences of *therefore* are used in this position. *However* and *on the contrary* stand between these two extremes, with 26.3% and 50.5% of sentence-initial uses, respectively. Finally, in a contrastive perspective, Altenberg's (2006) corpus-based study of CA placement in English and Swedish provides "clear evidence of individual positional 'profiles'" (ibid.: 15) among the set of 27 English CAs analysed in his corpus:

> Some [connectors] have a strong tendency to occur at E [end] (*though, anyway, then*), others prefer M [medial] (*accordingly, therefore, still*), while others are never placed there in the corpus (*furthermore, consequently, moreover*). Some connectors are extremely mobile without any clear preference for any position (e.g. *however, for example, for instance*), while others are practically restricted to one position (e.g. *besides* at I [initial]).

A similar picture emerges from his analysis of Swedish CAs. For example, Altenberg (ibid.: 30) shows that, although in total 29% of the CAs in his Swedish subcorpus are found initially, the percentage of sentence-initial uses per CA ranges from 0% (*nämligen*) to 87% (*därtill*). Likewise, depending on the conjunctive adjunct, the proportion of medial uses varies between 13% (*därtill*) and 89% (*således*). It must be stressed, however, that studies paying attention to lexical variation in placement remain the exception rather than the rule: to this day, most work on CA placement has tended to make fairly general statements concerning the positional patterns of conjunctive adjuncts for languages as wholes, making little or no reference to the internal variation characterising each language system. When lexis-sensitive comments are in fact available, they tend to be restricted to a few well-known cases, such as the tendency of *still* to occur initially, or the preference of *though* and *anyway* for sentence-final position (e.g. Biber et al. 1999: 891–892; Quirk et al. 1985: 643).

Finally, some – yet more infrequent – research has drawn attention to the influence of register variation on conjunctive adjunct placement. In an early study, Greenbaum (1969: 78–80) observed that, in a genre-varied corpus of about 100,000 words, two thirds (i.e. 51/75) of the medial CAs occurred in scientific writing, while a majority of the final CAs came from conversation. Like Jacobson's (1964) study, however, Greenbaum's analysis identifies a promising line of inquiry with respect to CA placement, but it is fairly limited empirically: it is based on 415 CA occurrences in total, spread across eight different registers. Arguably, the number of CA tokens per register is therefore too small to allow for any definite conclusions to be drawn regarding the influence of register on the position of conjunctive adjuncts (e.g. 22 CA tokens in the fiction texts, 16 occurrences in the newspapers, and only 9 in the spoken news reports). Greenbaum's findings were later corroborated on a larger scale by Biber et al. (1999: 890–892), who compared the placement patterns of English CAs in academic prose and conversation (see also Conrad 1999: 12–15 for a discussion of these results). They found that, although the initial position is the most frequent slot in both registers (viz. 55% of the occurrences in conversation, and 50% in the academic texts), conversation and academic prose display significant differences in their use of both medial (c. 40% in academic writing, against less than 2.5% in conversation) and final positions (c. 40% in conversation, against c. 10% in academic prose). Apart from these two studies, however, research into the influence of register variation on CA placement remains very limited. In addition, Greenbaum's (1969) and Biber et al.'s (1999) studies mostly focus on the differences between modes (viz. speech vs writing), rather than assessing the possible impact of register variation within a given mode (e.g. newspaper writing vs academic prose). Finally, I am not aware of any study that has systematically investigated either (i) the influence of register on conjunctive adjunct placement in a contrastive perspective; or (ii) the possible interaction

between lexical and stylistic factors with respect to CA placement. Thus, while Greenbaum's and Biber et al.'s studies undeniably draw attention to an important factor of influence, there is still much scope for research investigating the influence of register variation on CA placement in greater depth.

The objective of this chapter is to analyse and compare the placement patterns of English and French conjunctive adjuncts of contrast, while also assessing the influence of both lexis and register on the phenomenon. The position of conjunctive adjuncts of contrast is investigated in two written registers, viz. newspaper editorials, and academic writing. More specifically, the chapter sets out to answer the following four research questions:

i.   Do English and French differ in the positions that they allow for *and/or* prefer as regards conjunctive adjuncts of contrast? In this respect, reference will be made to the distinction made by Vinay and Darbelnet (1995: 15–16) between 'servitudes', i.e. the complex of "unalterable facts of the linguistic system" (ibid.: 15) and 'options', viz. the choices that users of a language tend to make within the limits imposed upon them by the servitudes of that language system.
ii.   To what extent does register influence (a) the placement patterns of conjunctive adjuncts within each language system; and (b) the differences between English and French in that respect (i.e. are the cross-linguistic differences stable across registers, or register-dependent)?
iii. To what extent does lexis influence the positional patterns of English and French conjunctive adjuncts?
iv.   If both lexis and register play a part, (how) do these two factors interrelate?

Thus, in addition to identifying differences between English and French patterns of conjunctive adjunct placement, the analyses carried out in this chapter will also attempt to clarify the role played by lexical and register factors on the phenomenon since, as explained above, the study of these variables remains rather sporadic in the current literature.

Since research into the influence of lexis and register remains rather marginal at this stage, no clear hypotheses could be formulated for research questions two to four, other than the fact that both lexis and register are likely to have an influence on the placement choices made by English and French writers. Based on the statements made on CA placement in the English and the French monolingual literature, on the other hand, it was possible to formulate research hypotheses for the first research question, which is concerned with the differences between English and French patterns of CA placement. Extensive research has been carried out on the positional patterns of English CAs. The phenomenon has been discussed not only in descriptive grammars (see e.g. Quirk et al. 1985: 643; Leech & Svartvik 2002: 139), but also in corpus-based diachronic linguistics (e.g. Lenker 2010; 2011; 2014),

corpus-based contrastive linguistics (e.g. Altenberg 1998; Kunz & Lapshinova-Koltunski 2014) or learner corpus research, which has often resorted to native data as a benchmark for assessing the degree of idiomaticity of learners of English with respect to CA use (e.g. Field & Yip 1992; Granger & Tyson 1996; Altenberg & Tapper 1998; Tankó 2004). As already discussed in Chapter 4 (Section 4.2.2), a majority of these studies have led to the convergent conclusion that English conjunctive adjuncts are used predominantly in initial position, which has been described as the "unmarked" (e.g. Biber et al. 1999: 891; Lenker 2014: 20), "normal" (Quirk et al. 1985: 643; Altenberg 1998: 122), "default" (Lenker 2011: sec. 4.2.2; Lenker 2014: 25) or "natural" (e.g. Halliday & Matthiessen 2004: 83; Altenberg 2006: 11) position for English CAs.

Unlike their English counterparts, French CAs have rarely been the object of rigorous, large-scale corpus-based investigations. Therefore, the placement tendencies outlined in the French literature largely remain to be verified empirically. They nevertheless provide useful insights for the formulation of hypotheses concerning differences between English and French CA placement patterns. Compared to English, the statements on CA placement made in the French literature are less categorical. While some researchers, such as Csüry (2006: 111), state that French CAs "have a marked tendency to occur at the beginning of the sentence (or clause)" [my translation], most scholars stress the mobility of French conjunctive adjuncts in the sentence (Rubattel 1982: 50; Hawkins & Towell 2001: 124), while also insisting that many conjunctive adjuncts occur in medial positions, and more particularly within the verb phrase. According to Grevisse and Goosse (2011: 1211), for example, "[a]dverbs that signal a logical relation are most frequently placed between the auxiliary and the participle" [my translation]. In fact, in a pilot study based on a sample of eleven newspaper editorials, Rossette (2009: 28) found that non-thematised (viz. non-initial) conjunctive adjuncts are about twice as frequent in French as in English.

Another element which may prove relevant for the formulation of hypotheses as regards CA placement in English and French pertains to more general differences in terms of word order flexibility between the two languages. Despite the fact that the two languages both follow the same broad 'Subject – Verb – Object' (SVO) word order principles, French has often been claimed to have a relatively flexible word order as compared to English, which tends to remain very faithful to the canonical SVO structure. Perhaps the most striking – and widely discussed – example of this difference concerns the markedly more common use of non-canonical syntactic structures circumventing the basic SVO word order – such as cleft, pseudo-cleft, topicalised or dislocated structures – in French as compared to English (see e.g. Trévise 1986; Lambrecht 2010 for a comparison of English and French spoken data; Carter-Thomas 2002; Bourgoin 2017 for contrastive studies based on written data). Another manifestation of this difference in syntactic

flexibility is the fact that French generally shows less resistance than English to the insertion of non-obligatory, parenthetical elements (including adjuncts) within the core 'Subject-Verb-Object' sequence of the clause (Chuquet & Paillard 1987: 155–156; 2017: 124; Guillemin-Flescher 1981: 126). In English, on the other hand, such interruptions of the canonical structure of the clause are usually not tolerated. This difference is exemplified in (4), where the temporal adjuncts used parenthetically between the verb and its direct object in French are both moved to final position in the English translation.

(4)   FR. Occuper, <u>pendant huit mois, en 1944–45</u>, un poste…
       EN. He held Cabinet rank <u>for eight months in 1944–45</u>….
                                                    (Chuquet & Paillard 1987: 156)

In view of all these elements, two main hypotheses can be formulated on the differences between English and French patterns of CA placement. Firstly, in terms of servitudes, it can be postulated that, because of its higher degree of syntactic flexibility, French might offer a larger number of syntactic slots for conjunctive adjuncts than English. Secondly, with respect to the preferences of each language system, it is expected that (i) initial conjunctive adjuncts will be more frequent in English than in French; while (ii) CAs in sentence-medial positions will be more common in French than in English.

The present chapter is divided into four main parts. Section 8.2 presents the Systemic Functional classification of CA positions adopted in this study. Section 8.3 provides a general overview of conjunctive adjunct placement across languages and registers, with a view to (i) identifying possible differences in CA placement between English and French; and (ii) assessing whether these cross-linguistic differences are stable across registers, or rather tend to be dependent on the communicative situation. Based on the results emerging from the register comparison of CA placement, Section 8.4 shows how conjunctive adjunct placement constitutes another example of a linguistic phenomenon situated at the interface between syntax and discourse, with different placement choices usually corresponding to distinct rhetorical effects at discourse level. Finally, Section 8.5 focuses on the influence of lexis on CA placement, with a view to assessing the degree to which the set of conjunctive adjuncts within each language tend to behave in roughly the same way, or rather display individual positional preferences. In a second stage, Section 8.5 also looks at the combined influence of lexical and register variation on English and French CA placement. Generally speaking, the discussion of quantitative results in this chapter proceeds as follows: the results are first discussed in a fully descriptive fashion. Then, the respective influence of the factors included in the analysis are tested statistically using the multifactorial method of Classification and Regression Trees (CART).

8.2   A Systemic Functional approach to conjunctive adjunct placement

In this chapter, I intend to describe the placement patterns of English and French conjunctive adjuncts of contrast by relying on the Systemic Functional system of thematic structure, i.e. the SFL system which is concerned with the linear ordering of constituents in the clause, and revolves around the notions of Theme and Rheme. The problem is that, whereas the Theme has been described fairly extensively in the SFL literature, little research has been devoted to the Rheme, which has usually been considered as one single entity (see Chapter 4, Section 4.2.1 for a detailed discussion). Yet, in order to provide a reliable contrastive description of CA placement in English and French, it is necessary to distinguish between the various positional options which are available within the Rheme – especially as medial slots have been described as important positions for French conjunctive adjuncts in the literature. The classification used in the present chapter therefore identifies five main positions for conjunctive adjuncts within the clause (or clause complex), i.e. two within the Theme, and three within the Rheme. The two thematic positions were identified on the basis of the distinctions within the Theme made in the Systemic Functional literature (cf. the distinction between textual, interpersonal and topical Themes). Due to the paucity of research on the Rheme, however, it was not possible to apply such ready-made SFL criteria to identify relevant positions within the Rheme. Whenever possible, inspiration was sought in the rare studies which have paid more attention to the Rheme (see e.g. Taglicht's (1984) 'marked rheme', or Morel & Danon-Boileau's (1998) 'post-rheme', discussed in Chapter 4), but not all of them were helpful (e.g. Morel & Danon-Boileau's (1998) 'post-rheme' is an intonationally-defined notion, and is therefore fully specific to spoken language; Fries's (1994) notion of N-Rheme only pertains to *experiential* constituents, and is therefore not suitable for describing the position of conjunctive adjuncts). As a result, the distinctions made within the Rheme are to a large extent corpus-driven: based on the patterns observed in the corpus data, I identified the positions which appeared to have functional relevance (see Section 8.4 for a discussion of the functions associated with each position) – while also drawing some inspiration from previous, non-SFL classifications of adjunct placement (e.g. Greenbaum 1969; Quirk et al. 1985; Hasselgård 2010b).

Before presenting the five-fold categorisation of adjunct positions, it is necessary to say a word about the unit of analysis chosen here. As explained by Fries (1995: 13), thematic structure can be analysed at several linguistic levels, including the phrase, the clause, the clause complex or even the paragraph. In this chapter, conjunctive adjunct placement is analysed with respect to the so-called 'independent conjoinable clause complex', also referred to as 'T-Unit'. In studies of thematic structure, the

T-Unit is defined as "an independent clause together with all hypotactically related clauses which are dependent on it" (Fries 1994: 229). For each T-Unit, one single topical Theme is identified. This means that, in an example such as (5), the fronted hypotactic clause is considered as the Theme of the entire clause complex. In an example such as (6), on the other hand, where the two clauses are related through parataxis, each clause receives its own analysis in terms of thematic structure. In the remainder of this section, the five positions are defined and exemplified.

(5) After the police arrived	I brought them to this cottage. (Thompson 2014: 160)
THEME	RHEME

(6) He	'd been a medieval history student in college	and I	was interested in medieval literature, too (Halliday & Matthiessen 2004: 406).
THEME 1	RHEME 1	THEME 2 (TEXTUAL + TOPICAL)	RHEME 2

The first category, labelled thematic 1, includes all the conjunctive adjuncts that occur either (i) at the very beginning of the T-Unit, as in (7) and (8); or (ii) immediately after another textual Theme, as in (9) and (10). The thematic 1 position thus corresponds to that which has been described as dominant in most accounts of English conjunctive adjunct placement, whether Systemic Functional or otherwise. In addition, CAs occurring at the beginning of T-Units with an ellipted subject, as in (11) and (12), are also included in the thematic 1 category (see e.g. Hasselgård 2010b: 44 for a similar decision concerning adjunct placement). As they occur before the first experiential element of the T-Unit, all the thematic 1 CAs perform the function of textual Theme.

(7)   Taken together, Studies 1 and 2 offer strong reason to believe that self-knowledge and social knowledge are mutually constraining. **However**, these findings are also consistent with another explanation.

(LOCRA-EN – Psychology)

(8)   Un individu construit son « identité pour soi » en se référant aux groupes auxquels il juge appartenir. **Cependant**, lorsqu'un individu appartient à un groupe du fait de son activité sociale et professionnelle, son identité diffère parfois de celle que ces membres lui attribuent.   (LOCRA-FR – Education)

(9)   All the polls suggest that voters do not support increasing the EU's power at the expense of our own, which is what the constitution does. So, **instead**, Labour is creating the illusion that there may be some kind of vote, possibly asking "Europe, in or out" - in order to try to take the sting out of the issue.

(Mult-Ed-EN – The Telegraph)

(10)   Pour exister sur la scène internationale, la Chine n'a plus aucun besoin de nous. <u>Et</u> **pourtant**, l'intérêt qu'elle porte à l'anniversaire de ses relations avec la France est un signe de sa volonté renouvelée d'ouverture sur le monde.
(Mult-Ed-FR – Le Figaro)

(11)   This political trope minimized class differences and **instead** focused on whiteness and the importance of hard work.
(LOCRA-EN – Anthropology)

(12)   Puisque ces deux notions ont été définies de façon purement procédurale, cela ne porte pas atteinte, mais **au contraire** confirme qu'il est possible d'avancer une justification purement procédurale du principe de majorité.
(LOCRA-FR – Political science)

The second type of thematic CAs, referred to as thematic 2, are placed right after an interpersonal Theme – i.e. after a clause-initial element which provides information on the stance of the writer towards his/her message – as in (13) and (14), where the interpersonal Themes are underlined. Like thematic 1 CAs, such markers occur before the first experiential element of the clause and are thus also considered to function as textual Themes. Interestingly, however, these uses contravene the allegedly typical order of Themes, where textual Themes are supposed to occur before interpersonal Themes (see Halliday & Matthiessen 2004: 79; see also Chapter 4 for more on this).

(13)   This variable again showed no relationship to impressions of trustworthiness from the targets' faces […]. <u>Interestingly</u>, **however**, we did observe a relationship between the amount of time that targets cheated and the targets' self-reports of their past cheating behavior. (reverse scored)
(LOCRA-EN – Psychology)

(14)   If there must be trumpets, at least let them be played well. To learn, children should, it is said, start at seven, and the very thing has now popped up to entice them into taking a dedicated interested in the instrument: the so-called pTrumpet. It is made of plastic. Plastic sounds bad, but in this case is good. Best of all it sounds just like a brazen trumpet. <u>Most usefully</u>, **though**, it costs only £100: far less than an easily dented traditional instrument.
(Mult-Ed-EN – The Telegraph)

The first rhematic category, labelled rhematic 1, includes the conjunctive adjuncts which occur between the topical Theme (whether marked or unmarked) and the main verb phrase of the T-Unit. Examples of rhematic 1 CAs are provided in (15) to (18), where the topical Themes are underlined. Examples (15) and (17) illustrate rhematic 1 CAs used after an unmarked Theme, whereas in (16) and (18), the Theme preceding the CA is marked (with a spatial and a temporal adjunct, respectively, functioning as topical Theme). In both cases, rhematic 1 uses of

CAs are akin to what Hannay and Gómez-González (2012) refer to as 'thematic parentheticals', where a parenthetical element (which may, but must not be a conjunctive adjunct) occurs directly after the Theme (see Chapter 4 for more on thematic parentheticals).

(15) Japan, the Philippines and others worry that the Obama administration is lukewarm in upholding US treaty commitments to them. <u>China</u>, **on the other hand**, fears the US is trying to contain it.

<div align="right">(Mult-Ed-EN – The Financial Times)</div>

(16) In Maryland, one of the "bluest" states in the nation, the push for accountability came out of a desire of those on the left at the state level to take control over, and increase funding for, what state actors saw as failing high-poverty schools. <u>In Utah</u>, **by contrast**, the movement for standards-based reform was largely championed by the political right.

<div align="right">(LOCRA-EN – Education)</div>

(17) Parce qu'elles ont été principalement engagées dans les activités caritatives et philanthropiques, les femmes sont plutôt associées à des enjeux tels que le bien-être des enfants, la vie morale des familles, etc. <u>Les hommes</u>, **en revanche**, ont été historiquement investis sur des enjeux qui excèdent le cadre de la cellule familiale : défense de la natalité, de l'enseignement catholique…

<div align="right">(LOCRA-FR – Political science)</div>

(18) Certes, la décision de s'expatrier a toujours procédé d'un malaise, économique ou politique, mais l'attrait du lieu de destination a joué un rôle dans les mouvements migratoires dont ont bénéficié les divers pays d'Amérique, à commencer par les Etats-Unis, ou, en Europe, la France. <u>Depuis les années 1970</u>, **au contraire**, les inégalités économiques et la pauvreté des pays du Sud sont la principale cause des tentatives de certains de leurs habitants pour se frayer un chemin jusqu'aux pays du Nord.

<div align="right">(Mult-Ed-FR – Le Monde)</div>

The rhematic 2 category encompasses all the conjunctive adjuncts occurring within the predicate of the T-Unit. It is a fairly broad category which includes cases of CAs occurring (i) within the main verb phrase – as in (19) and (20); (ii) between the verb and its complement(s) – as in (21) and (22); or (iii) more rarely, within the (sequence of) complement(s), as in (23) and (24).

(19) It would let Labour off far too lightly to pretend that the Dodo's verdict on the caucus-race in Alice's Adventures in Wonderland - "Everybody has won and all must have prizes" - applies to the local election results. But the outcome of Thursday's voting <u>has</u> **nevertheless** <u>produced</u> a set of looking-glass conclusions.

<div align="right">(Mult-Ed-EN – The Guardian)</div>

(20)   La littérature spécialisée a depuis longtemps identifié isolément chacun des trois termes (globalisation, intégration européenne, discours expert) de ce type de discours. Le présent article <u>a</u> **cependant** <u>montré</u> que la crise offre un terrain particulièrement propice à leur utilisation conjointe.
(LOCRA-FR – Political science)

(21)   The government could have chosen to wait until the end of the investigation. Provided the FCA's recommendations had teeth, this could have led to greater competition and lower prices. Mr Osborne <u>opted</u> **instead** <u>for a more radical shake-up</u>.   (Mult-Ed-EN – The Financial Times)

(22)   Bien que systématiquement relancés sur leurs souvenirs ayant trait à la municipalité et à ses actions, rares ont été les interviewés à développer des interprétations explicites du rôle de la municipalité de Nanterre. Parmi les quelques narrations recueillies, on <u>retrouve</u> **cependant** <u>la même oscillation entre intégration et exclusion</u>.   (LOCRA-FR – Anthropology)

(23)   This latter requirement supports the maximized - or, on the basis of proceeding remarks, an optimized - satisfaction of whatever purpose, self-interested or not, an action sets out to achieve. The means-ends efficiency axiom of rationalist accounts is <u>conditional</u>, **however**, <u>on the factual specification of the ends of the action in question at the time that the means are chosen</u>.   (LOCRA-EN – Sociology)

(24)   De récents travaux sur la comparaison ont rappelé l'importance de l'échelle nationale comme cadre de résolution des problèmes publics. Nous sommes <u>attentives</u> **cependant** <u>à ne pas tomber dans le nationalisme méthodologique</u>.
(LOCRA-FR – Political science)

Thus, the rhematic 2 position is rather similar to one of the positions identified by Halliday (1985: 81) in the first edition of his *Introduction to Functional Grammar*, where he had drawn attention to a tendency of non-thematic conjunctive adjuncts to occur at the boundary between the Mood (viz. the subject along with the part of the verbal group that expresses tense or modality) and the Residue (viz. the rest of the clause). The rhematic 2 category is nevertheless slightly broader, so as to accommodate cases such as (25) and (26), where the CA is placed later than the Mood (ending with *does* and *ont*, respectively), within the Residue. In addition, CAs occurring within (or between) complements, such as (23) and (24) above, are also part of the Residue, rather than occurring at the Mood-Residue boundary.

(25)   Athletics companies like Nike don't produce apparel anymore; they produce a brand, that assortment of ideas and associations that tie together the various commodities from which they extract revenue (Lury 2004). This does not mean, **however**, that real physical labor is a thing of the past.
(LOCRA-EN - Anthropology)

(26)   Les études visant à tester cette hypothèse ont montré que la charge en mémoire de travail affectait peu ou pas les effets de congruence (Kessler & Meiran, 2010; Kiesel, Wendt, & Peters, 2007; Meiran & Kessler, 2008). Cohen-Kdoshai et Meiran (2009) ont constaté **en revanche** que lors des premiers essais dans une tâche, l'effet de congruence mesuré était sensible à la charge en mémoire de travail.   (LOCRA-FR - Psychology)

Finally, rhematic 3 conjunctive adjuncts occur after all the verb complements in a given T-Unit. In many cases, such as (27) and (28), this position corresponds to the sentence-final position.[2]

(27)   The latest poll, therefore, represents an extraordinary turnaround in just a few weeks and could hardly have come at a more critical moment in a debate that has been rumbling on for 40 years or more. Several caveats need to be entered, **however**.   (Mult-Ed-EN – The Telegraph)

(28)   En renonçant au plaisir d'apparaître dominateur, Sarkozy, victorieux de lui-même, a maximisé ses chances de succès. Dans la bataille du débat, Ségolène Royal est loin de tout perdre **cependant**. (Mult-Ed-FR – Le Figaro)

However, the rhematic 3 category also includes conjunctive adjuncts which occur after all the complements of the main verb in the T-Unit but are still followed by one or several optional constituents. Typically, these constituents are adjuncts with scope over the entire T-Unit (instead of simply modifying the verb), as in (29), where the the temporal adjunct following the CA could easily be fronted – or even removed (see e.g. Hasselgård 2010b: 48–53 on the distinction between sentential and predicational adverbials). However, other elements may also follow rhematic 3 adjuncts. For example, rhematic 3 CAs can be followed by a right-dislocated constituent co-referring with a pronoun placed earlier in the clause, as in (30), where *les gosses* refers to the same entity at the subject pronoun *ils.*[3]

(29)   As is evident in the following interview excerpt, Moriah was well aware of this self-confidence and spoke of it in an effortless, nonchalant way […]. This confidence was quickly diminished, **however**, within weeks of starting college.   (LOCRA-EN – Education)

---

2.   Despite the fact that these CAs are the last elements in the clause, they do not correspond to Fries's (1994) N-Rheme, which refers to the last *experiential* constituent in the clause (see Chapter 4 for more details).

3.   Note that in (30), *les gosses* could possibly be viewed as a written equivalent to Morel and Danon-Boileau's 'post-rheme', which was defined as an optional intonational constituent produced with a flat intonation, a low pitch and a reduced intensity at the end of an oral paragraph, and could take the form of a nominal group co-referring with a pronoun expressed earlier in the clause (as in: *Mais elle est vieille sa filleule*; see Chapter 4 for more details).

(30)   Elle serait à qui veut la prendre, mais qui la veut ? Les gosses que, comme
        lundi, peu de profs encadrent. Ils sont bien, **pourtant**, <u>les gosses</u>.

(Mult-Ed-FR – Libération)

So far, the description of the adjunct classification has focused on the most com-
mon and straightforward types of patterns encountered in the corpus data. As
is often the case in authentic data, however, a number of cases did not fit neatly
within these five categories. One type of structure, in particular, deserves to be
discussed in some detail – especially as it occurred quite frequently in the cor-
pus. It pertains to conjunctive adjuncts used within or right after an impersonal
clause occurring at the beginning of the T-Unit. This pattern is exemplified
in (31) to (35), where the impersonal clauses are underlined. Some Systemic
Functional linguists – including Halliday himself – would consider the Theme
in such examples to simply be the impersonal pronoun (viz. English *it, there* or
French *il*), with the rest of the clause functioning as Rheme (see e.g. Halliday &
Matthiessen 2004: 98). However, it is arguably problematic to assign the label of
topical Theme to impersonal pronouns: as explained by Thompson (2014: 165),
such pronouns have no true representational function, "and therefore [do] not
fulfil the thematic criterion of expressing experiential meaning" (ibid.). In fact,
the issue of how to analyse impersonal structures is all the more important in
the framework of the present English-French contrastive study as, according to
Rossette (2009: 21), impersonal pronouns are involved in a much larger number
of constructions in French than in English (e.g. *il faut, il s'agit de, il y a, il semble
que*, etc.). As a result, the stance taken towards the thematic analysis of imper-
sonal pronouns may have serious implications for the outcome of the corpus
study.

(31)   The long-term consequences of developmental influences are now thought
        to be mediated mainly through epigenetic changes, whereby molecular
        structures attached to DNA and affecting its expression are altered (Fall
        2011). <u>It is clear</u>, **however**, <u>that</u> the effect of being small in early life is
        exacerbated by adiposity in later life.   (LOCRA-EN – Anthropology)

(32)   Dans les déclarations concernant des détériorations cognitives, on retrouve
        à première vue (Tableau 4) des régularités déjà observées précédemment.
        Chez les ouvriers, on mentionne davantage les symptômes concrets
        entraînés par la déficience (oubli des médicaments, difficultés pour
        s'occuper des papiers). Du côté des personnes les plus favorisées, les
        déclarations font la part belle au thème du déclin des « facultés »
        (« altération », « baisse »). <u>Il est</u> **cependant** <u>surprenant que</u> ce soit chez les
        ouvriers que le terme « Alzheimer » soit surreprésenté.

(LOCRA-FR – Sociology)

(33)   Si les enseignants ont parfois recours à la force envers leurs élèves, les punitions corporelles telles qu'on les entend, sont rares. <u>Il s'agit</u> **plutôt** <u>de</u> gestes impulsifs, sous l'emprise de la colère ou pour mettre un terme à un comportement jugé dangereux.   (LOCRA-FR – Education)

(34)   Bref, on imputera aux tares de la société américaine la flambée de violence de Ferguson. <u>Il y a</u> **pourtant** une leçon plus générale à tirer de ces événements tragique.   (Mult-Ed-FR – Libération)

(35)   This had been the option favoured by the Government, a policy strongly supported by Andrew Lansley, the former Health Secretary, who was shuffled out of that post in 2012. <u>There are</u>, **however**, problems with it.
          (Mult-Ed-EN – The Independent)

As it is indeed questionable to set a referentially empty word as the point of departure of the message, impersonal constructions as wholes were analysed as Theme when they occurred at the beginning of a T-Unit. More specifically, depending on the semantic content of the impersonal clause, two main types of analyses were performed. In some cases, the *it*-clause appeared to convey a predominantly interpersonal meaning, while the constituent that followed was the element which truly carried experiential meaning. In Examples (31) and (32), for instance, the meanings put forward by the *it*-clauses are roughly equivalent to that of comment adjuncts (or 'disjuncts') such as *clearly* and *surprisingly*, respectively (see e.g. Larsson 2017a; 2017b on the functional equivalence between *it*-clauses and *-ly* adverbs). Accordingly, such preposed clauses have sometimes been referred to as 'interpersonal grammatical metaphors', to account for the fact that instead of being expressed in the form of comment adjuncts as is typically the case, "interpersonal meanings are experientialized and treated as if they were 'content' meanings" (Thompson 2014: 168; see also Thompson & Hunston 2008: 47–51; Herriman & Bostrom Aronsson 2009). Thus, following Herriman and Bostrom Aronsson (2009) or Thompson (2014: 168–171), such impersonal clauses as the ones in (31) and (32) are analysed as interpersonal elements in a multiple Theme, with the following experiential element (e.g. *the effect of being small in early life)* being considered as the topical Theme of the T-Unit, and functioning as the true point of departure of the message. Accordingly, CAs occurring after (or within) such constructions are labelled thematic 2.[4]

     Alongside impersonal clauses conveying a stance meaning, the corpus data also contained a fair number of impersonal constructions which did not convey

---

4.   As opposed to these authors, however, I *only* consider impersonal constructions (i.e. introduced by the pronouns *it, there* in English; or *il* in French) conveying meanings of modality as interpersonal Themes. While for Thompson (2014) or Herriman and Bostrom Aronsson (2009), initial projection clauses such as *I think, I believe, I suspect*, etc. are also considered to play the role of interpersonal Theme, I take the view that the first-person subjects in such constructions should be regarded as topical Themes since, unlike impersonal pronouns, they refer to concrete entities in the outside world (viz. the writers themselves).

any interpersonal meaning. Such structures were mostly found in French, which displayed a fairly large frequency of CAs used within or after impersonal clauses such as *il y a, il existe, il s'agit*, etc., as in (33) and (34). Because these constructions do not express stance, it was not possible to analyse them as interpersonal Themes, in the same way as structures such as *it is clear that* or *il est suprenant que*. Rather, following Rossette (2009: 21) in her study of French thematic structure, such constructions were analysed as the topical Theme of the clauses which they introduced. In other words, in (33), the sequence *il s'agit de* is analysed as Theme, and in (34), the Theme corresponds to *il y a* (see also Thompson 2014: 166 for an equivalent analysis of clauses introduced by an existential *there*). As a result, the conjunctive adjuncts in Examples (33) to (35) are labelled rhematic 1, since they occur directly after the topical Theme of the T-Unit.

A summary table of the classification adopted in the present chapter is provided in Table 24, where English and French examples are provided for each of the five positions.

Table 24. Summary table of the five adjunct positions

		EXAMPLES	
POSITION	**DESCRIPTION**	**ENGLISH**	**FRENCH**
THEMATIC 1	At the beginning of the T-Unit *or* after another textual Theme	**However**, these findings are also consistent with another explanation.	**Cependant**, lorsqu'un individu appartient à un groupe du fait de son activité sociale et professionnelle, son identité diffère parfois de celle que ces membres lui attribuent.
THEMATIC 2	After an interpersonal Theme	<u>Interestingly</u>, **however**, we did observe a relationship between the amount of time that targets cheated and the targets' self-reports of their past cheating behavior.	<u>Il est</u> **cependant** <u>surprenant que</u> ce soit chez les ouvriers que le terme « Alzheimer » soit surreprésenté.
RHEMATIC 1	Between the topical Theme and the verb phrase	<u>China</u>, **on the other hand**, fears the US is trying to contain it.	<u>Les hommes</u>, **en revanche**, ont été historiquement investis sur des enjeux qui excèdent le cadre de la cellule familiale.
RHEMATIC 2	Within the predicate (verb phrase + complements)	But the outcome of Thursday's voting <u>has</u> **nevertheless** <u>produced</u> a set of looking-glass conclusions.	Le présent article <u>a</u> **cependant** <u>montré</u> que la crise offre un terrain particulièrement propice à leur utilisation conjointe.
RHEMATIC 3	After all the verb complements	Several caveats need to be entered, **however**.	Dans la bataille du débat, Ségolène Royal est loin de tout perdre **cependant**.

Based on this classification, the position of French and English conjunctive adjuncts of contrast was analysed in both the Mult-Ed corpus of quality newspaper editorials and the LOCRA corpus of research articles, so as to assess the impact of register variation on CA placement. The corpus study focused on conjunctive adjuncts used in finite main clauses, so as to ensure maximum comparability of the analyses in the two languages (e.g. avoiding that the results be influenced by the higher ratio of CAs in verbless or hypotactic clauses in French than in English). In addition, only the conjunctive adjuncts occurring with a raw frequency of at least 50 occurrences in *each* subcorpus were examined, in order to ensure that the number of occurrences per CM in each position would be sufficiently high to allow for generalisations to be made on the individual behaviour of the various CAs. The English CA *yet* and the French CA *or* were excluded from the analysis, as both markers are fixed in clause-initial position.[5] In total, eight English and seven French CA types (viz. English *by contrast, however, instead, nevertheless, nonetheless, on the other hand, still, though;* French *au contraire, cependant, en revanche, néanmoins, plutôt, pourtant, toutefois*), corresponding to 5,940 and 4,488 tokens, respectively, were coded for position. This amounts to c. 80% of the total number of English CAs, and c. 75% of the French conjunctive adjuncts in the corpus (excluding *yet* and *or*).

8.3   Conjunctive adjunct placement across languages and registers: A general overview

This section presents a broad overview of the placement patterns of conjunctive adjuncts of contrast in English and French. It starts with brief intralingual descriptions of the placement preferences of English and French (taken separately) in the two written registers, viz. academic prose and editorials (Section 8.3.1). Section 8.3.2 then provides a cross-linguistic comparison of English and French CA placement patterns in the two registers. Finally, based on the outcome of the cross-register and cross-linguistic comparisons carried out in the previous sections, Section 8.3.3 attempts to evaluate the respective influence of language and register on conjunctive adjunct placement, by applying the statistical method of Classification and Regression Trees (CART) on the data set.

---

5.   English *yet* and French *or* are hybrid adverbial markers, which share features with both the category of adverbs (e.g. a capacity to combine with coordinators) and coordinators (e.g. an obligatory clause-initial position).

8.3.1   Conjunctive adjunct placement in English and French: Intralingual cross-register comparisons

This section aims to provide a general description of English and French conjunctive adjunct placement in two written registers, viz. research articles and newspaper editorials. The objective is to assess the degree to which CA placement appears to be affected by register variation *within* each language system – a phenomenon which has received little attention in the literature, where fairly general principles of CA placement tend to be formulated for languages as wholes. The proportion of English and French conjunctive adjuncts of contrast in each of the five positions per register are provided in Tables 25 and 26, respectively. They are represented graphically in Figures 25 and 26.

Table 25. Placement patterns of English conjunctive adjuncts in LOCRA and Mult-Ed (in percent)

|	LOCRA	Mult-Ed
Thematic 1	**64.8%**	**38%**
Thematic 2	1.7%	2.7%
Rhematic 1	23.1%	**39.8%**
Rhematic 2	7.3%	13.2%
Rhematic 3	3.2%	6.3%
Total	100% (3,407)	100% (2,533)
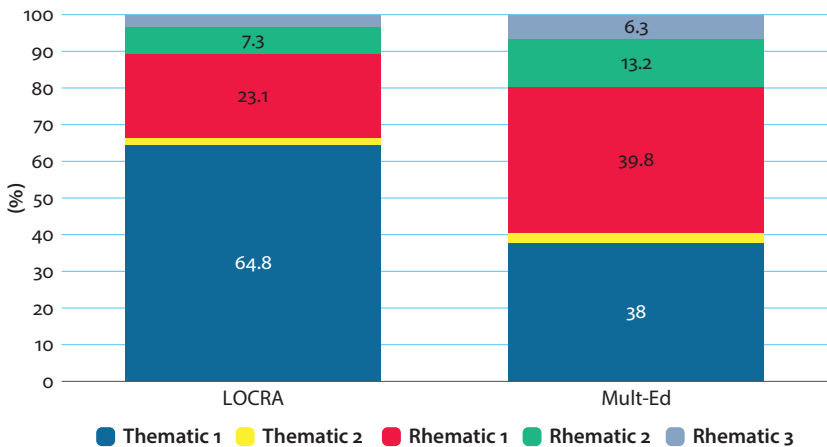
Figure 25. Placement patterns of English conjunctive adjuncts in LOCRA and Mult-Ed (in percent)

Figure 25 shows that register variation has an appreciable effect on conjunctive adjunct placement in English written language. The most striking result in this respect is the fact that the preferred positions for conjunctive adjuncts of contrast differ in the two registers. In academic prose, CAs are used predominantly in thematic 1 position, as in (36). Such results are in perfect line with the statements found in the literature, where the sentence-initial position is nearly unanimously described as the "natural", "normal" or "default" position for English conjunctive adjuncts. In the editorials, by contrast, the thematic 1 position can hardly be said to constitute the default position for conjunctive adjuncts, as it does not even account for half of the occurrences (38%). Instead, the CA tokens are roughly equally distributed between the thematic 1 and the rhematic 1 positions (with each position accounting for c. 40% of the occurrences), as in (37) and (38), respectively. More generally, rhematic CAs appear to be markedly more frequent in the editorials than in the research articles: whereas in LOCRA, the two thematic positions represent as much as two thirds of the occurrences (viz. 66.5%), in Mult-Ed, a majority of CA tokens are placed after the topical Theme, in the rhematic part of the T-Unit (59.3%), with each rhematic category being about twice as populated in the editorial as in the academic register. Therefore, these results show that the statements found in the literature, which insist on the key importance of the initial position for English conjunctive adjuncts, need to be qualified to some extent: while they appear to hold true in some communicative situations (e.g. academic prose), they are not valid across the board. In other contexts, such as newspaper editorials, rhematic positions are also important, or even predominant. It is nevertheless worth noting that the high proportion of rhematic 1 CAs in the editorials may be partly specific to the type of CAs analysed here, as both Altenberg (2006: 16) and Lenker (2011) have noted that CAs of contrast are among the most common markers to occur after a subject or a fronted adjunct.

(36)   Almost all of the couples reported more costs than benefits. **However**, most couples reported they would continue off-shifting and that overall, despite the numerous costs, off-shifting was the best arrangement for them.
(LOCRA-EN – Sociology)

(37)   There were local reasons why things went so badly wrong, and local actors who made disastrous choices. And there were some points where this long intervention might have taken a different turn, with conceivably better results. **Nevertheless**, rarely in modern history can military force have been exerted over such an extended period to such little purpose.
(Mult-Ed-EN – The Guardian)

(38)   Countries such as China and India have lost interest in international negotiations and are going it alone to improve domestic energy efficiency. These steps, **however**, are no substitute for global action.
(Mult-Ed-EN – The Financial Times)

As stands out from Figure 26, the placement patterns of French conjunctive adjuncts also vary across registers. As was the case in the English data, the editorial and the academic registers differ with respect to the position that they tend to prefer for conjunctive adjuncts of contrast.

Table 26. Placement patterns of French conjunctive adjuncts in LOCRA and Mult-Ed (in percent)

	LOCRA	MULT-ED
THEMATIC 1	**52.8%**	31.7%
THEMATIC 2	0.6%	1.4%
RHEMATIC 1	8.7%	16.9%
RHEMATIC 2	37.3%	**47.5%**
RHEMATIC 3	0.7%	2.5%
Total	100% (2,738)	100% (1,750)
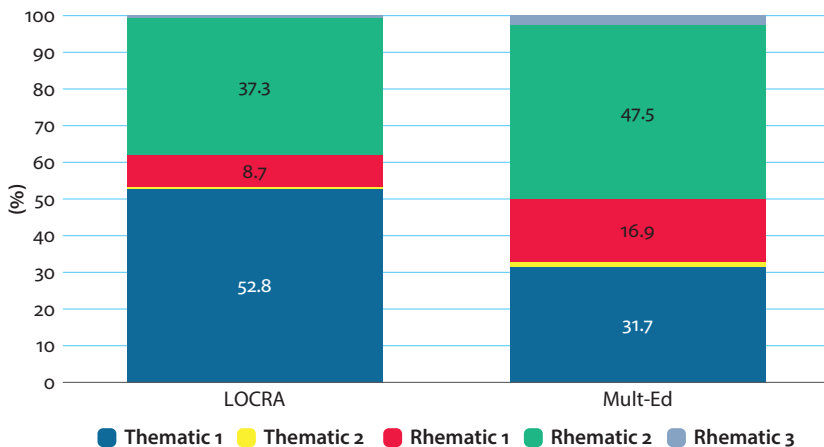
Figure 26. Placement patterns of French conjunctive adjuncts in LOCRA and Mult-Ed (in percent)

In LOCRA, the most frequent position is thematic 1 – as in (39), which accounts for just over half of the CA tokens in this subcorpus. In Mult-Ed, on the other hand, the most frequent slot is the rhematic 2 position, where the CA is placed within the predicate, as in (40). In addition, rhematic 1 CAs – as in Example (41) – are also about twice as common in the editorials as in the research articles. Similarly to English, the ratio of thematic and rhematic positions thus differs between the registers: in academic prose, CA tokens are roughly equally distributed between

the thematic and the rhematic subparts of the T-Unit. The distribution is more uneven in the editorials, where two thirds of the CAs are used rhematically. It must nevertheless be underlined that, in spite of the preference of academic prose for thematic 1 CAs, and editorials for rhematic 2 markers, both positions are common choices in both registers. In this, the results provide empirical confirmation of the statements made in the French literature, where the tendency of CAs to occur both initially and medially, and more particularly within the verb phrase, has been underlined (see e.g. Grevisse & Goosse 2011: 1211; Csüry 2006: 111).

(39)   Que ces objets soient laissés sur place ou remportés à la suite du dépôt ne change rien au fait qu'ils ne peuvent être considérés comme des offrandes. **Au contraire**, ils remplissent une fonction de captation et attestent que les transferts matériels se font toujours dans deux directions.

(LOCRA-FR – Anthropology)

(40)   La dette, qui a dépassé la barre symbolique de 1 000 milliards d'euros (63 % du PIB), serait en conséquence sur une pente qui la conduira à 67 % du PIB en 2005, 10 points de plus qu'en 2001. M. Sarkozy a **néanmoins** promis à ses collègues ministres des finances réunis en Irlande de tenir l'objectif de 3 % en 2005.   (Mult-Ed-FR – Le Monde)

(41)   La panne sèche menace dans de nombreux secteurs, notamment dans l'industrie et le secteur public local, dont les investissements sont pourtant essentiels au retour de la croissance. En France **pourtant**, jamais les banques n'ont autant sélectionné les risques, au point de déserter des pans entiers de l'économie.   (Mult-Ed-FR – Le Monde)

In summary, the corpus data shows that depending on the communicative situation, English and French writers alike tend to make partly diverging positional choices for conjunctive adjuncts of contrast. One important implication of this observation is that it calls into question the relevance of formulating general rules of placement for languages as wholes, as is typically done in the literature. For example, whereas it is certainly the case that writers of research articles in English should be encouraged to opt for thematic 1 uses of conjunctive adjuncts, an editorial in which most CAs would occur initially might appear slightly unnatural in view of the conventions of the register. This is especially true since, as will be shown later in this chapter, choices regarding CA placement are not without significance, but can play a true role with respect to the communicative goals of a given text type (see Section 8.4).

8.3.2   Cross-linguistic comparison of English and French conjunctive adjunct placement: A register-sensitive account

As established in the previous section, register clearly has an impact on CA placement within each language system. This section now provides a cross-linguistic

comparison of the placement choices made by English and French writers in both academic and editorial texts. The objective is (i) to identify possible differences between languages with respect to both their possible and preferred positions for conjunctive adjuncts of contrast, but also (ii) to assess whether the influence of register uncovered within each language system also has an impact on these cross-linguistic differences. The proportions of use of conjunctive adjuncts in each position per language and per register are provided in Table 27. The most frequent positions in each subcorpus are in bold. For each register, green cells indicate that a given position is more common in English than in French, whereas the positions that are more common in French than in English appear in orange.

Table 27. Placement patterns of English and French conjunctive adjuncts in LOCRA and Mult-Ed (in percent)

	LOCRA		MULT-ED	
	ENGLISH	FRENCH	ENGLISH	FRENCH
THEMATIC 1	**64.8%**	**52.8%**	**38%**	31.7%
THEMATIC 2	1.7%	0.6%	2.7%	1.4%
RHEMATIC 1	23.1%	8.7%	**39.8%**	16.9%
RHEMATIC 2	7.3%	37.3%	13.2%	**47.5%**
RHEMATIC 3	3.2%	0.7%	6.3%	2.5%
TOTAL	3,407 (100%)	2,738 (100%)	2,533 (100%)	1,750 (100%)

The first observation to be made on this table is that in both registers, all five positions are attested in both English and French (albeit sometimes rather rarely, as in the case of the thematic 2 and rhematic 3 positions). In other words, in terms of 'servitudes', viz. inherent possibilities offered by the linguistic systems (see Vinay & Darbelnet 1995), no language seems to make more positions available for conjunctive adjuncts than the other. Rather, the two languages appear to have a similar set of possibilities at their disposal regarding CA placement. This stands in opposition with one of our hypotheses, which predicted that in view of its supposedly greater degree of syntactic flexibility, French might offer a wider range of slots than English.[6]

---

6.   Admittedly, the classification used in the present study may be considered rather broad, since it 'only' identifies five different positions within the T-Unit. It is of course possible that a more refined classification may have uncovered differences in possible positions between the two languages. Faced with this eventuality, I decided to add another coding layer to the data, by making further distinctions within each of the three rhematic categories. Within the

Whereas the results emerging from the corpus analysis do not reveal any cross-linguistic differences in terms of servitudes, they do point to clear differences in frequencies of use of the different positions. In Mult-Ed, for instance, the two languages differ in terms of the positions which they most often choose for conjunctive adjuncts of contrast. In the French editorials, the most frequent slot is the rhematic 2 position, as in (42), where the CA occurs within the main verb phrase of the T-Unit. While this position hosts about half of the French CAs of contrast in Mult-Ed, it is not very frequent in the English subcorpus, where it accounts for only about one CA token out of eight (13.2%).

(42)   Si l'on en croit les témoignages réunis par les journalistes, leurs accusations sont clairement fondées. M. Poutine les a **cependant** d'office rejetées en déclarant cette élection "la plus propre de l'histoire de la Russie".

(Mult-Ed-FR – Le Monde)

The English editorials, on the other hand, display not one but two preferred positions for CAs, viz. the rhematic 1 position (39.8% of the occurrences) and the thematic 1 slot (38% of the occurrences), which are exemplified in (43) and (44), respectively.

(43)   The first group will vote for Ukip or - this is a prospect that often goes ignored - not vote at all. Chasing them is a fool's errand. The second group, **however**, can be tempted by the kind of tangible improvements in their lives that can only be delivered by practical, level-headed government.

(Mult-Ed-EN – The Financial Times)

(44)   They are critical of M Aristide's rule, in particular his use of thugs to intimidate political opponents. **However**, they recognise that he was duly elected four years ago and that his immediate departure could create a power vacuum inviting even greater chaos than at present.

(Mult-Ed-EN – The Telegraph)

In LOCRA, by contrast, the two languages display the same preferred position for conjunctive adjuncts: in the English as well as the French research articles, the most frequent category is the thematic 1 slot, as exemplified in (45) and (46).

---

rhematic 1 category, I distinguished between CAs occurring (i) after an unmarked Theme; or (ii) after a marked Theme. Within the rhematic 2 category, I made a distinction between CAs used (i) within the verb phrase; (ii) between the verb phrase and its complement(s); and (iii) within or between the complement(s). Finally, I differentiated between rhematic 3 CAs (i) used at the very end of the T-Unit; and (ii) followed by (an)other optional constituent(s). Even from a more fine-grained classification such as this one, however, all these slots were well attested in both languages. Such results corroborate the conclusion that there seems to be no marked difference in the possibilities offered by each language with respect to CA placement.

(45)   As Sue (2010) points out, if microaggressions are recognized as problematic, then reactions such as anger are "understandable and normative" (p. 58). **However**, if a microaggression is not recognized as such and is considered acceptable, then any reaction to the act can be seen as "pathological" (p. 58).

(LOCRA-EN – Education)

(46)   Un joueur ne peut jouer deux fois consécutivement la balle. **Cependant**, si le joueur A effectue la première touche et que son partenaire (joueur B) touche la balle en deuxième touche, le joueur A qui est intervenu en premier peut toucher une seconde fois la balle dans le total des trois touches allouées (joueur A, joueur B, joueur A).   (LOCRA-FR – Education)

Despite this point of commonality, however, Table 27 does reveal marked cross-linguistic differences in frequency of use of (some of) the positions in the research articles. In fact, as is clear from the 'colour pattern' in Table 27, the type of frequency differences emerging from the corpus data are strikingly similar across registers: in both LOCRA and Mult-Ed, French displays a markedly higher frequency of rhematic 2 CAs than English, whereas all the other positions – and, more strikingly, the thematic 1 and the rhematic 1 positions – are more common in English than in French.

These results have two implications. Firstly, they clearly demonstrate that within a common range of servitudes (or possible positions), English and French differ in the positional options that they tend to prefer for conjunctive adjuncts of contrast (in terms of frequencies of use). In order to provide a comprehensive account of differences in CA placement between the two languages, it is therefore essential to think not only in terms of what is possible (i.e. grammatically acceptable), but also in terms of what is more or less likely to occur in each language. Secondly, the fact that the same tendencies emerge from the two subcorpora strongly suggests that with respect to conjunctive adjunct placement, cross-linguistic differences between English and French are relatively stable across registers. In other words, although register seems to exert a significant influence on the placement choices made by writers within each language system (see Section 8.3.1), it does not affect the type of differences uncovered between English and French. This in turn suggests that the preferences of each language system (i.e. the inclination of French to use CAs in rhematic 2 position, and the tendency of English to place them in thematic 1 and rhematic 1 positions) may influence CA placement to a larger extent than the communicative situation in which the markers are used. Interestingly, such findings are consistent with the results obtained from the comparison of frequencies of CAs in Chapter 7 (Section 7.4), where conjunctive adjuncts of contrast were shown to be significantly more frequent in English than in French, whatever the register analysed. Further reflections on the respective influence of language and register on conjunctive adjunct placement are found in

the next section. In the remainder of this section, I focus more specifically on the frequency differences between languages (which, for the most part, are therefore roughly identical in both registers).

In the introduction to this chapter, it was hypothesised that thematic 1 conjunctive adjuncts would be more common in English than in French, which would display a larger proportion of sentence-medial CAs, and more particularly a marked tendency to place these items within the verb phrase (i.e. in rhematic 2 position). The first part of this hypothesis is supported by the corpus evidence: thematic 1 CAs are indeed more typical of English than French, in both registers. It is nonetheless worth noting that, since thematic 1 CAs are also quite popular in French, differences in frequency of use of this position are perhaps less striking than for some other categories (e.g. in Mult-Ed, English only displays about 6% more of thematic 1 CAs than French, which also resorts to this position in about a third of the cases). The second part of the hypothesis, on the other hand, which predicted that medial CAs would be more common in French than in English, needs to be qualified. It is true that the preference of French for CAs in rhematic 2 position is largely confirmed by the corpus results: those CAs are 3.5 times as frequent in French as in English in Mult-Ed, while this ratio rises to 5 to 1 in the academic subcorpus. However, the importance of rhematic 1 CAs in the English subcorpus (which are about twice as frequent as their French equivalents in both registers) clearly shows that medial positions in general are not the prerogative of French.[7] Instead, it is more accurate to state that English and French tend to prefer different rhematic positions for conjunctive adjuncts: while in English, CAs which are not used thematically tend to occur between the topical Theme and the verb phrase – as in (47) – in French non-thematic conjunctive adjuncts are most often placed within the predicate, as in (48). Importantly, the fact that English and French appear to prefer different rhematic positions within the T-Unit provides ample justification for the

---

7. Interestingly, between c. 30% (in Mult-Ed) and c. 50% (in LOCRA) of the French rhematic 1 CAs occur within or after either: (i) an impersonal clause functioning as Theme, as in: (1) _Il existe_ **toutefois** _un scepticisme plus radical qui consiste à priver de légitimité l'idée même de démocratisation, qui ne serait pas un désir de partage du bien de tous, mais la volonté sournoise d'imposer à tous le goût de quelques-uns._ (LOCRA-FR – Political science); or (ii) a cleft sentence – referred to as 'predicated Theme' by Halliday and Matthiessen (2004: 95–98) or Thompson (2014: 155–156) – as in: (2) _C'est_ **pourtant** _l'inverse qui_ s'est produit dans les urnes (Mult-Ed-FR – Le Monde). In English, on the other hand, these uses account for less than 5% of the rhematic 1 CAs. Thus, in addition to displaying markedly different frequencies of rhematic 1 CAs, English and French also display qualitative differences in their use of this position.

decision to make distinctions within the Rheme: had the Rheme been considered as a single, undifferentiated unit (as is typically the case in the SFL literature), this cross-linguistic difference in frequency of use of the different medial positions would have gone completely unnoticed.

(47)   For weeks the assumption in some western capitals had been that Mr Putin's next step would be to send Russian troops into eastern Ukraine, coming to the "rescue" of pro-Moscow separatists who demand independence from Kiev. <u>On Wednesday</u>, **however**, he performed an apparent U-turn, raising hopes of a diplomatic solution.   (Mult-Ed-EN – The Financial Times)

(48)   Ce sont des combats en costumes « militaires », parfois scénarisés, qui visent à représenter des champs de batailles de l'époque. Les pratiquants soulignent **cependant** la distance qui existe entre ces activités et la réalité historique.   (LOCRA-FR – Anthropology)

The differences in use of medial positions are probably traceable to a more general contrast between the English and French language systems with respect to adjunct (and more specifically, adverb) placement. According to White (1991) or Osborne (2008), one fundamental difference between the English and French systems of word order is that, whereas French allows for adverbs to be used between a verb and its object, as in (49), this pattern is most often prohibited in English. In contrast, English allows for adverbs to occur between the subject and the verb of the sentence, as in (50), which would arguably be forbidden in French.

(49)   Jean <u>mange</u> **souvent** de la glace [*John <u>eats</u> **often** ice cream].
                                                            (Osborne 2008: 127)

(50)   John **often** <u>eats</u> ice cream [?Jean **souvent** <u>mange</u> de la glace].   (ibid.)

As the corpus data shows, at least regarding the category of conjunctive adjuncts (as opposed to frequency and circumstantial adverbs and adjuncts, for instance), it is probably a bit too strong to speak in terms of 'allowed' versus 'prohibited' patterns of use in such cases, since the corpus data displays both instances of (i) French adjuncts used between the subject and the verb – as in (51) and (52) – and (ii) English adjuncts used between a verb and its object – as in (53) and (54). However, what could more accurately be referred to as 'strong systemic preferences' for one or the other pattern may explain the higher proportion of rhematic 1 CAs in English, and rhematic 2 CAs in French.

(51)   C'est injuste pour la Pologne. Seul pays de l'Union européenne à ne pas avoir connu de récession depuis 2008, elle vient de traverser deux décennies de transition, de modernisation et d'ouverture remarquables. Au sein de l'UE, elle a acquis un poids inédit. <u>L'Ukraine</u>, **en revanche**, a déçu et irrité les dirigeants européens.   (Mult-Ed-FR – Le Monde)

(52)   Certes, cette évolution duale pourrait au contraire être perçue comme
       un enrichissement, et avoir en effet positif sur l'élargissement du champ
       d'application empirique des différentes variantes de QCA. <u>Le danger</u>,
       toutefois, résiderait dans le manque de cumulativité de l'ensemble, et
       plus encore dans la raréfaction d'espaces d'innovation communs où les «
       développeurs » de QCA pourraient continuer à affiner l'approche.
       (LOCRA-FR – Political science)

(53)   The British are a tolerant people, and our ability to absorb millions of
       immigrants over the past two generations does us credit. This liberality
       <u>depends</u>, **however**, <u>on an unspoken bargain between the citizen and the
       state</u>.   (Mult-Ed-EN – The Telegraph)

(54)   In contrast, Kalwij (2010) considers that most of this impact may be
       spurious and find no significant impact of family expenditures per children
       neither on the probability to have children nor on completed family size.
       His analysis did not <u>consider</u>, **however**, <u>the net transfers received by families
       through the tax and benefit system</u>.   (LOCRA-EN – Sociology)

Importantly, it must also be stressed that the frequent use of English CAs in rhematic 1 position only partly contradicts the assumption that English makes it difficult to interrupt the SVO sequence of the clause or sentence by means of optional parenthetical elements (see e.g. Chuquet & Paillard 2017: 155; Guillemin-Flescher 1981: 126). The rhematic 1 position encompasses CAs used not only (i) after an unmarked topical Theme, i.e. most typically a subject, as in (55), but also (ii) after a marked topical Theme, which most frequently takes the form of an adjunct, as in (56). Only in the former case does the conjunctive adjunct interrupt the dependency structure of the T-Unit.[8] As reflected by the addition of a secondary layer of coding to the rhematic 1 CAs, distinguishing between the CAs introduced by a unmarked Theme, and those following a marked Theme, over 40% of the rhematic 1 CAs in English occur after a marked topical Theme. Thus, close to half of the rhematic 1 CAs in English still occur outside the SVO sequence of the T-Unit.

(55)   Even if Mr Xi is genuine about tackling corruption, the suspicion is his
       allies will not be touched. <u>His adversaries</u>, **on the other hand**, are far
       from safe.   (Mult-Ed-EN – The Financial Times)

---

8.   It must be specified that marked Themes also include instances where a constituent which would normally function as verb complement or object is used as Theme, as in: *Crystal clear, by contrast, is her closeness to the President* (Mult-Ed-EN – The Telegraph). In such cases, the CA is in fact included within the dependency structure of the clause. However, such cases only constitute a minority of the marked Themes uncovered in the corpus data.

(56)   A month ago, the ICM poll taken in the wake of Maria Miller's cabinet resignation seemed to show little Tory benefit from the improving economy. <u>A month on</u>, **however**, it begins to look as though the voters are more confident about the economy, as one might expect.

(Mult-Ed-EN – The Guardian)

All in all, in spite of the predominance of the rhematic 1 position in English, the total proportion of CAs occurring within the core SVO structure of the sentence thus remains higher in French than in English, which still displays a greater propensity to use CAs of contrast outside the dependency structure of the T-Unit. On the other hand, the corpus results certainly contrast with Rossette's (2009: 28) preliminary study, which had shown that non-thematic CAs were twice as frequent in French as in English. Table 28 presents the results which are obtained when grouping the three rhematic positions in each language. The table makes clear that the difference is much more limited than argued by Rossette: depending on the register, French rhematic CAs are no more than 1.4 times (in LOCRA) or 1.1 times (in Mult-Ed) as frequent as their English counterparts. The discrepancy between her results and my own is all the more striking as her study was based on a sample of newspaper editorials, which is precisely the register where the cross-linguistic difference in proportion of rhematic CAs is the smaller. The difference between the two studies is most probably due to the very small size of Rossette's corpus, which consisted of only eleven newspaper editorials – as opposed to the c. 8,500 editorials analysed in the present book.

Table 28. Proportion of rhematic CAs in English and French per register

	ENGLISH	FRENCH	RATIO OF THE DIFFERENCE
LOCRA	33.6%	46.7%	1.4
MULT-ED	59.3%	66.9%	1.1

Finally, the two languages resemble each other in their infrequent use of thematic 2 and rhematic 3 positions – although these both appear to be more frequent in English than in French. The low frequency of these positions is probably unsurprising. On the one hand, the low frequency of thematic 2 CAs is in line with Hasselgård's (2004b: 72) corpus-based analysis of multiple Themes in English, which revealed that overall, multiple Themes containing both interpersonal *and* textual elements are relatively rare (viz. they represent less than 7% of the multiple Themes in her corpus). In addition, according to Halliday and Matthiessen (2004: 79), the most natural ordering of elements in the multiple Theme is one in which textual units precede interpersonal ones. This means that thematic 2 uses disrupt the typical order of elements in the multiple Theme, which could also explain the low frequency of

CAs directly following an interpersonal Theme in the corpus. That being said, it is noteworthy that for a number of examples of thematic 2 CAs in the corpus, it would have sounded rather unnatural to place the interpersonal element after the conjunctive adjunct. Compare, for example, (57) with (58); or (59) with (60). This may underline a need for the refinement of the SFL rule regarding the sequencing of elements in the Theme: in practice, this order seems to partly depend on the type of textual and/or interpersonal elements being used in the Theme.

(57) Scenario studies suggest that a weak state leaving the eurozone might see its currency devalue up to 60 per cent and its GDP halved (Belke, 2011, p. 9). <u>Importantly</u>, **however**, the costs of euro exit would not be limited to the exiting country alone. (LOCRA-EN – Political science)

(58) Scenario studies suggest that a weak state leaving the eurozone might see its currency devalue up to 60 per cent and its GDP halved (Belke, 2011, p. 9). **However**, <u>importantly</u>, the costs of euro exit would not be limited to the exiting country alone. (adapted version of Example (57))

(59) Yes, there is a sizeable constituency of non-believers (particularly among the young), but the majority retain a relationship with faith. <u>Worryingly</u>, **however**, our survey also found that 48 per cent of the public believe Christians are afforded less protection than members of other faiths. (Mult-Ed-EN – The Telegraph)

(60) Yes, there is a sizeable constituency of non-believers (particularly among the young), but the majority retain a relationship with faith. **However**, <u>worryingly</u>, our survey also found that 48 per cent of the public believe Christians are afforded less protection than members of other faiths (adapted version of Example (59)).

The generally low frequency of rhematic 3 CAs in the corpus, on the other hand, seems logical in view of the type of texts analysed in the present study. A number of researchers have underlined that CAs in final position are relatively common in speech (see e.g. Biber et al. 1999: 891; Lenker 2010: 200; Haselow 2011, 2012; Degand 2014). According to Haselow (2012: 187), one of the main reasons why speakers use CAs (which he calls 'particles') in final position pertains to the unplanned nature of their linguistic productions:

> [The use of final particles] derives from the interactive nature of language production in conversations, where the different contributions have not been pre-planned and are added cumulatively. Final particles are used strategically to add information on the semantic relation between two units, each with its own illocutionary force, *after* they have been produced, without previous planning of this relation […]. In this sense, the use of final particles is an efficient strategy to create cohesion between two propositions in the utterance-chaining style that is characteristic for spoken discourse, where planning is restricted and complex

> relations often evolve or are recognized only during the production of an utter-
> ance itself […]. The use of an utterance with a final particle is a case of delayed
> linking, which reflects the preference for right-tending, coordinative supplemen-
> tation under conditions of real-time planning […].

By contrast, neither of the registers investigated in the present chapter is subject
to similar temporal constraints to the ones that characterise spontaneous spoken
communication. Writers of both newspaper editorials and research articles have
the opportunity not only to plan but also to edit their texts, which enables them to
signal a given relation between two discourse units earlier than at the end of the
complete cohesive sequence (see Conrad & Biber 2000: 66 for a similar observation
on stance adverbs). Cross-linguistically, the higher frequency of rhematic 3 CAs in
English as compared to French may be a consequence of the reluctance of French
to use grammatical words at the end of a sentence. In their contrastive account of
English and French – and more particularly in the section devoted to word order
and thematic structure – Vinay and Darbelnet (1995: 226) explain that "[f]rom a
stylistic point of view the final position in French is preserved for content words".
This explanation is all the more plausible as, within the set of rhematic 3 CAs
(which correspond to the CAs occurring after all the elements that are part of the
dependency structure of the T-Unit), the proportion of markers occurring at the
very end of the T-Unit – as in (61) and (62) – is strikingly higher in English than
in French (viz. 76% vs 55% in Mult-Ed, and 57% vs no more than 6% in LOCRA).
In French, on the other hand, most rhematic 3 CAs tend to be followed by optional
elements, as in (63). Also note that the higher proportion of final CAs in the edito-
rial as compared to the academic register in the two languages may be a result of
the more speech-like, informal tone of this register (see e.g. Hundt & Mair 1999;
Westin 2002; see also Chapter 7 on this).

(61)   Jones and Baumgartner (2005) found highly concentrated attention in the
       low-partisanship US media context. A desire to appeal to rival partisan
       readerships might create different dynamics in Spanish news coverage,
       however.   (LOCRA-EN – Political science)

(62)   Hier, après avoir rappelé «l'urgence», «la nécessité» et «la légitimité» d'une
       intervention militaire, François Hollande a affirmé que «la France ne
       sera pas seule». Beaucoup en doutent encore, y compris sur les bancs du
       Parlement. Ils se trompent **pourtant**.   (Mult-Ed-FR – Libération)

(63)   L'organisation interne, aussi bien du Parlement européen que de la
       Commission et du Conseil, marquée par la pratique de la négociation et du
       compromis entre des acteurs ressortissant presque toujours aux trois partis
       dominants (PPE, PSE, ALDE), laisse peu de latitude aux opposants qui
       n'entreraient pas dans ce jeu. L'action et la collaboration européennes des
       opposants se développent **néanmoins** depuis les années 1990.
                                                           (LOCRA-FR – Political science)

Finally, with respect to the thematic 2 category, it is perhaps worth noting that a greater diversity of patterns was found in the English than in the French sub-corpus. In English, thematic 2 CAs were roughly equally distributed between (i) CAs used after a comment adjunct, as in (64); and (ii) CAs used after an inter-personal *it*-clause functioning as interpersonal grammatical metaphor, as in (65). In French, on the other hand, the former pattern was not attested. Rather, all the cases of thematic 2 CAs observed in the French data correspond to conjunctive adjuncts used within or after an interpersonal *it*-clause expressing the stance of the writer, as in (66).

(64) 'United in Diversity' - thus ran the motto that the abandoned EU Constitutional Treaty envisaged for the European Union (EU). It underlines that integration in the Union takes place with due respect for the diversity among its members. <u>Clearly</u>, **however**, there are tensions between unity and diversity. (LOCRA-EN – Political science)

(65) The particular target of Dr Sentamu's anger is racism, which is hardly surprising: as Bishop of Stepney, he was searched by police eight times in six years. <u>It is a shame</u>, **however**, <u>that</u> he has invested so heavily in the concept of "institutional racism". (Mult-Ed-EN – The Telegraph)

(66) Cependant, parce que peu d'information est disponible dans les études sur le nombre de mois passés en institution et l'âge de l'enfant à l'arrivée en institution, il est difficile de bien saisir la signification de ce résultat. <u>Il demeure clair</u>, **néanmoins**, <u>que</u> la possibilité de développer un attachement sécurisant est mince. (LOCRA-FR – Psychology)

This difference may reveal a very subtle discrepancy in servitudes between English and French, namely the impossibility of using conjunctive adjuncts after a sentence-initial comment adjunct in French, whereas this pattern is possible in English. Alternatively, this may possibly be ascribed to the very small number of thematic 2 CAs in the French subpart of the corpus (viz. 39 occurrences in total). The analysis of a larger data set may have uncovered such uses in French as well.

In summary, this section has demonstrated that, while English and French do not appear to display any marked differences in the positional possibilities that they offer for conjunctive adjuncts of contrast, they display clear differences with respect to the positions that they tend to prefer for these markers. Such results have important implications, notably for translation and translator training. As pointed out by Loock (2018: 115–116), although issues pertaining to word order are crucial for the idiomaticity or naturalness of a translation, it also constitutes one of the areas in which machine translation is still inadequate. Such questions thus remain the prerogative of human translators, i.e. one of the aspects over which they still have a true added value as compared to automated translation tools. It is

therefore crucial that apprentice translators be made aware of the subtleties of the target language, and the main differences between the source and target systems with respect to word order.

8.3.3   Respective weight of language and register on conjunctive adjunct placement

So far, the analyses carried out in this chapter have demonstrated both (i) an influence of register variation on CA placement within each language system (see Section 8.3.1); and (ii) substantial cross-linguistic differences in the positional preferences of English and French with respect to CA placement, in each of the two registers investigated. When combining these two factors of influence, on the other hand, it appeared that the differences between languages did not seem to be affected by register variation. As a reminder, Table 29 summarises the main results of the cross-linguistic comparison of conjunctive adjunct placement per register. As appears clearly from this table, the differences between English and French are highly stable across communicative situations: in both registers, all five positions were attested in the two languages. In addition, in both the editorials and the research articles, rhematic 2 CAs were more typical of French, whereas all the other CAs were found to be more frequent in English. In fact, the only difference between the two registers is that in the research articles, the preferred position was the same in both languages (viz. thematic 1), whereas in the editorials, each language had its own preferred position(s) (viz. rhematic 1 and thematic 1 in English, and rhematic 2 in French).

Table 29. Cross-register comparison of English and French conjunctive adjunct placement patterns

	Mult-Ed	LOCRA
Number of possible positions for CAs	EN = FR	EN = FR
Frequency of thematic 1 CAs	EN > FR	EN > FR
Frequency of thematic 2 CAs	EN > FR	EN > FR
Frequency of rhematic 1 CAs	EN > FR	EN > FR
Frequency of rhematic 2 CAs	FR > EN	FR > EN
Frequency of rhematic 3 CAs	EN > FR	EN > FR
Same preferred position in English and French	No (FR = rhematic 2; EN = rhematic 1 + thematic 1)	Yes (thematic 1)

Based on these observations, I formulated the preliminary conclusion that, even though both register and language play a significant role in determining the

position of CAs of contrast, the preferences of each language system influence CA placement to a greater extent than the register in which the conjunctive adjuncts are used. This section attempts to establish more clearly the respective influence that language and register have on conjunctive adjunct placement in English and French, by applying the statistical method of Classification and Regression Trees (CART) on the corpus data (see Breiman et al. 1984; Berk 2008: chap. 3).

As explained in Chapter 5, CART trees are an inferential statistical method that attempts to predict a given outcome (here: the position of conjunctive adjuncts) from a set of independent variables or 'predictors' (here: language and register), based on the (linguistic) data fed into the model. The main principle along which the CART method operates is that, within a given data set, it attempts to find the factor which acts as the best predictor of the outcome at hand (viz. the position of CAs, in this particular case). This best predictor gives rise to a first binary distinction within the data set, which can be read as the first stage in a decision tree (e.g. if the language is English, I am likely to observe outcome X; if the language is French, I am likely to observe outcome Y). The procedure is then repeated on each subset of the data: the algorithm keeps identifying the next best predictor, until it reaches a point where additional divisions based on these predictors would no longer result in a more effective prediction of the outcome. In sum, the most important element to remember here is that the higher a given predictor appears on the tree, the greater its role in predicting the outcome (viz. the more influential it is). The factors leading to lower divisions are also significant, but they play a more limited role than those occurring higher in the tree. The label and the colour in each node (or 'leaf') of the tree indicate the category which represents the majority of the data points in this node. The five figures which occur below the position labels in each node indicate the percentage of the data in that node which belongs to each category (e.g. in the top node of the tree, which represents the whole data set: 50% of the CAs occur in thematic 1, 2% in thematic 2, 22% in rhematic 1, 23% in rhematic 2, and 3% in rhematic 3), while the percentage value in each node indicates the proportion of the entire data set that is represented by this node.[9]

The CART tree predicting the position of conjunctive adjuncts as a function of language and register based on LOCRA and Mult-Ed is provided in Figure 27. The tree provides strong statistical support in favour of the conclusion that language is more influential than register in determining positional choices with

---

9.   The trees reported in this chapter are always 'pruned', viz. they provide a predictive rather than a descriptive view of the phenomenon of CA placement (see Chapter 5 for more details). Non-pruned (i.e. descriptive) versions of all the CART trees presented in this chapter are provided in Appendix 3.

respect to CAs of contrast. The fact that language gives rise to the first division in the tree indicates that it is the best predictor of the position of conjunctive adjuncts of contrast in our corpus. Based on the corpus data, the model is able to make a clear distinction between the CA instances which occur in the French data and those which come from the English subcorpus – in spite of the fact that thematic 1 CAs represent the most frequent category in each data set. From the percentage values comprised in each node, we can assume that this first division was probably triggered mostly by the difference in frequency of the rhematic 1 (30% in English vs 12% in French) and rhematic 2 positions (10% in English vs 41% in French) in the two languages. This conclusion is supported when looking at the lower nodes in the tree, which show that each of these two rhematic positions is typical of one language in particular.
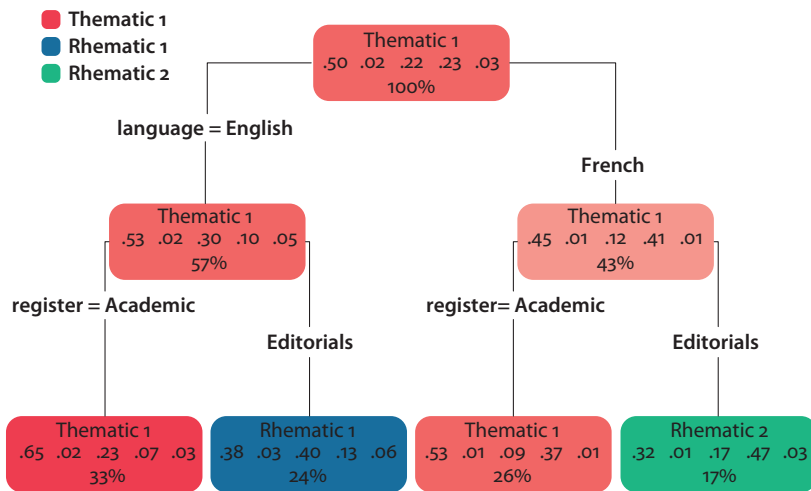
Figure 27. Classification tree for conjunctive adjunct placement as a function of language and register

In addition, the CART tree reveals that register also has a significant influence on CA placement, since it also gives rise to divisions in the model. However, the fact that this division only appears second in the tree indicates that register is secondary with respect to language. In fact, the tree clearly shows that register influences CA placement *within* each language system: within each language, register is the factor that will determine whether the CAs are used preferably in thematic 1 position – as in the research articles – or in rhematic position – as in the editorials – with each language tending to prefer a different rhematic position in the editorials (viz. rhematic 1 in English, and rhematic 2 in French). It is worth noting that, despite the fact that language influences CA placement to a larger extent

than register, the common preference of English and French for thematic 1 CAs in academic prose, and rhematic CAs in editorial writing points to a certain degree of cross-register similarity across languages. In other words, register acts on CA placement in strikingly similar ways in the two languages (see also Lefer 2009: 245 on prefixation in English and French for similar findings).

Interestingly, the results obtained here stand in opposition with previous register-sensitive contrastive studies on English and French, which demonstrated that the differences uncovered between the two languages were partly dependent on the type of communicative situation analysed. In a contrastive study of clause linking in English and French, for instance, Cosme (2006) attempts to test the hypothesis that interclausal coordination tends to be more frequent in English than in French, by comparing the frequency of interclausal *and* and interclausal *et* in a comparable corpus made up of fiction and news texts. She reports that, while the hypothesis is verified in the fiction subpart of the corpus, no significant cross-linguistic difference in frequency of interclausal coordinators is found in the news subcorpus. Cosme (ibid.: 81) thus concludes that "the difference in frequency between clausal coordination with *and* and clausal coordination with *et* may be genre-related". In a similar vein, Granger (2014) compares the frequency of English and French stems (viz. lexical bundles containing a subject and a verb, such as *I agree with* or *it is important that*) in a comparable corpus containing newspaper editorials and parliamentary debates, with a view to testing the hypothesis that French tends to be more formulaic than English. While her hypothesis is clearly confirmed in the editorial subcorpus, she also finds that stems are much more common in the English than in the French parliamentary debates, in terms of both types and tokens. Granger (ibid.: 64) thus concludes that "it is […] unwise to formulate general contrastive statements on lexical bundle frequency. As our figures show, different genres can display diametrically opposed tendencies". Similar results are also presented in Fløttum et al.'s (2006: 53–55) linguistic comparison of three academic disciplines (viz. business, linguistics and medicine) in English, French and Norwegian. Trying to tease out the respective influence of language and discipline on the use of a range of linguistic features used by researchers to refer to either their own or other researchers' voices in their writing (viz. metatext, bibliographical references, adversative conjunctions, negation, first-person subjects and indefinite pronoun subjects), Fløttum et al. (ibid.: 54) conclude the following:

> Both the discipline factor and the language factor have a significant effect for all the features. But for the majority of features, discipline turns out to be more important than language. That is, authors of research articles tend to write more like their disciplinary colleagues writing in other languages than like their language community co-members writing in other disciplines – with respect to most of the features that we have chosen to study here.

Finally, in her contrastive study of English and French discourse markers in eight spoken registers, Crible (2018: 85) suggests that register variation tends to have a larger effect on the frequency of discourse markers than cross-linguistic differences, with the two languages displaying a strikingly similar cross-register patterning regarding the use of discourse markers.

Therefore, when compared with previous register-sensitive contrastive research on English and French, the results obtained here seem to suggest that the respective weight of the linguistic and register factors may depend on the type of linguistic phenomenon investigated. Whereas for some linguistic phenomena (such as the frequency of use of interclausal coordinators or lexical bundles), cross-linguistic differences appear to play a secondary role with respect to register variation, in other cases (such as CA placement) the preferences of each language system trump the effect of register, resulting in cross-linguistic differences that are stable across registers. This hypothesis is supported by the fact that my results are – once more – in line with another recent contrastive study on cohesion, viz. Kunz and Lapshinova-Koltunski's (2015) comparison of the use of cohesive devices (i.e. reference, substitution and conjunction) across eight English and German written registers (viz. popular science, tourism leaflets, prepared speeches, political essays, fictional texts, corporate communication, instruction manuals and corporate websites). Based on a corpus-based analysis of the frequency and types of cohesive devices used in each language and register, the authors also found that "contrasts are more pronounced between the two languages English and German than between registers" (ibid.: 283).

Importantly, despite the fact that register is less influential than language when it comes to conjunctive adjunct placement, it is still worth wondering why this factor has such a strong impact on the placement choices made within each language system, with academic writing demonstrating a clear preference for thematic 1 CAs, while newspaper editorials very frequently resort to rhematic positions. In the following section, I argue that this marked difference in ratio of thematic vs rhematic CAs in the two registers can be ascribed – at least partly – to the rhetorical effects that may be achieved by CAs used in certain positions. I show that, similarly to the syntactic patterning of conjunctive markers of contrast described in Chapter 7, CA placement constitutes a good example of the syntax-discourse interface.

8.4   Conjunctive adjunct placement at the syntax-discourse interface

In his book devoted to focusing adverbs in English, Taglicht (1984: 12) distinguishes between two main kinds of rules governing the ordering of words and phrases in a sentence. The first rule is "purely syntactic, in the sense that sequence

is uniquely determined by syntactic dependency relations", whereas the second is "syntactic and contextual, which means that considerations of coherence and emphasis, in addition to syntax in a narrower sense, have a part to play" (ibid.). According to Taglicht, the purely syntactic rule makes it possible to discriminate grammatically correct structures from syntactically erroneous ones (e.g. *Do you hear?* as opposed to *Hear do you?*). By contrast, when two (or more) grammatically correct word orderings are possible (e.g. *I saw Mary* vs *Mary I saw*), choices will be governed by contextual as well as syntactic considerations (ibid.). In both English and French, most grammars and textbooks devoting a section to conjunctive adjuncts refer to the potential of these items for syntactic mobility: they highlight the co-existence of various grammatically correct positions for conjunctive adjuncts, thus focusing on the 'purely syntactic' side of CA placement. This aspect was also one of the foci of Section 8.3, which was partly concerned with identifying the possible positions for CAs of contrast in each language. As Lenker (2011) rightly points out, however, grammars and textbooks typically provide little or no information on the reasons why writers should select one position over the others for a given conjunctive adjunct – i.e. they do not provide clear guidelines concerning the 'contextual rules' that govern the choice of one position in particular within the range of grammatically correct options.

In this section, I attempt to pinpoint some of the main motivations that induce such placement choices. Zooming in on more limited samples of CA tokens spread across the two languages and registers, I try to demonstrate that the different positions available for conjunctive adjuncts within the T-Unit are not equivalent, but rather tend to create distinct effects at the discourse level (Sections 8.4.1). I argue that in this respect, CA placement can be viewed as another manifestation of the syntax-discourse interface, and that the discourse functions achieved by CAs used in certain positions may in turn explain the marked differences in frequencies of thematic and rhematic CAs observed between the editorial and the academic registers within each language system (Section 8.4.2). As compared to the results presented earlier, the present section offers a more 'qualitative' account of CA placement. The discussion focuses primarily on the three categories that are most frequent in the data set and represent over 90% of the conjunctive adjuncts in the corpus, viz. the thematic 1, rhematic 1 and rhematic 2 categories. These are also the positions which were identified by the CART method as giving rise to significant differences between the registers (see Section 8.3.3 above). The other two categories, viz. thematic 2 and rhematic 3, are only discussed briefly in this section. Importantly, the following discussion does not claim to account for *all* the contextual factors influencing CA placement. Rather, the objective is to underline some of the most general tendencies emerging from the corpus data. Finally, it also needs to be stressed that the analyses presented in this section build on

previous research by Altenberg (2006) and Lenker (2011; 2014), who also looked into the discourse functions fulfilled by CAs in some syntactic slots, and more particularly in medial positions.

8.4.1   Conjunctive adjunct placement in English and French: A range of discourse functions

8.4.1.1   *Thematic 1 conjunctive adjuncts as 'pure' markers of conjunction*

A close inspection of thematic 1 conjunctive adjuncts across languages and registers reveals that, when they are placed in thematic 1 position – as in Examples (67) to (70) – CAs tend to be restricted to their most basic function, viz. they express a logical link (here: contrast) between two segments of discourse. By selecting CAs as the main point of departure of their message, writers lay emphasis on the logical progression of their discourse, and their main objective is to ensure "that the connection between two clauses is clearly signalled as the reader […] moves from the first to the second connect" (Lenker 2010: 44). This is particularly clear from Examples (68) and (70), where the contrastive CAs are part of a larger sequence of discourse-structuring devices (underlined in the examples), with each marker introducing a different step in the writer's textual development. In these examples, sentence-initial CAs function as 'signposts' or 'beacons', as it were, explicitly pointing out to the reader the boundaries between different argumentative stages in the text.

(67)   Accurate estimation of information naturally leads to the establishment of mechanisms that enable intensive or sometimes even aggressive information searches. **By contrast**, overestimation of information is not likely to lead to a similar response because the longer the policy appears calibrated (Lichtenstein, Fischhoff and Lawrence, 1982, 307), the more confident the policy-makers become about the reliability of the information at their disposal.   (LOCRA-EN – Political Science)

(68)   Of course, the court is not faultless. Jonathan Sumption, a UK Supreme Court judge, argues that the ECHR's interpretation of the right to a private and family life has now gone far beyond the intentions of its postwar framers. **However**, even the most contentious ECHR judgments cannot ultimately count as more than minor irritations for the British government. And withdrawal from it might involve more complications than the government anticipates.   (Mult-Ed-EN – The Financial Time)

(69)   Tout le monde s'accorde à penser que l'école doit instruire, c'est-à-dire former les élèves à savoir ce qu'il en est du monde (énoncés descriptifs et constatifs). **En revanche**, que l'école doive éduquer est plutôt controversé, et sans doute en France plus qu'ailleurs pour des raisons historiques.
                                                                     (LOCRA-FR – Education)

(70)   Deuxième exemple, cette fois en creux : la hausse du Smic. <u>Assurément</u>, celle-ci ne peut être attribuée à Nicolas Sarkozy, qui semble plutôt avoir défendu une ligne orthodoxe. <u>Et certes</u>, il est tout à fait légitime de penser que le travail peu qualifié est trop peu rémunéré par les entreprises. **Toutefois**, imposer un salaire minimum peut conduire à l'effet pervers que de nombreux travailleurs se voient privés d'emploi, car les employeurs estiment que leur apport à l'entreprise serait inférieur au coût de les employer.   (Mult-Ed-FR – Libération)

In the remainder of this section, it will be demonstrated that the thematic 1 position is in fact the only syntactic slot in which CAs perform a strictly linking function. In all the other positions, conjunctive adjuncts tend to do more than 'simply' signal a logical connection between two discourse units: they are also at the service of other discourse considerations. More specifically, non-initial CAs tend to lay focus not (only) on the logical relationship itself, but (also) on other constituents in the T-Unit, with each position emphasising a different part or aspect of the message.

8.4.1.2   *Rhematic 1 conjunctive adjuncts*

Rhematic 1 conjunctive adjuncts, which occur between the topical Theme and the main verb phrase of the T-Unit, are exemplified in (71) to (74), where the topical Themes are underlined. A careful examination of the data set sheds light on a range of discourse effects performed by CAs in rhematic 1 position, some of which were remarkably stable throughout the corpus. It must be noted that, for the most part, the functions of English and French rhematic 1 CAs uncovered in the corpus data are consistent with those discussed by Altenberg (2006) and Lenker (2011; 2014) in their studies of medial CAs in English, which suggests that the discourse functions associated with this syntactic slot are partly constant across language.

(71)   Childless women, for the most part, have employment patterns that are much more similar to childless men's patterns. <u>Patterns for women with one or more children</u>, **however**, are distinctly different.
                                                  (LOCRA-EN – Sociology)

(72)   <u>In 1945</u>, though Britain had lost nearly 400,000 military dead, and another 65,000 civilians, the streets were deep lined with cheering crowds. <u>In 2003</u>, **by contrast**, there were few crowds in the streets as today's notabilities took their places - and certainly no cheering.   (Mult-Ed-EN – The Guardian)

(73)   Mais Max Weber ne se préoccupe pas de sa candidature. Conséquence : il est placé en position non éligible sur la liste de Heidelberg (Bade) qui ne correspond pas au canton où il devait initialement candidater. Max Weber n'est pas élu. <u>Sa femme</u>, **en revanche**, devient députée.
                                                  (LOCRA-FR – Political science)

(74) Au Pérou, les discours sur l'assistancialisme de Juntos se réfèrent au renforcement du caractère oisif et paresseux des pauvres. Juntos freinerait le « développement », car il habituerait les pauvres à le demeurer et à être des « assistés », dépendants dès lors de l'État. <u>En Bolivie</u>, **en revanche**, les Bonos seraient assistancialistes parce qu'ils ne luttent pas contre les causes structurelles de la pauvreté et participent alors à sa reproduction.

(LOCRA-FR – Political science)

(75) Childless women, for the most part, have employment patterns that are much more similar to childless men's patterns. **However**, <u>patterns for women with one or more children</u> are distinctly different

(adapted version of Example (71)).

The first and most widespread function of rhematic 1 conjunctive adjuncts of contrast emerging from the corpus data is to focus the reader's attention on the experiential content expressed in the topical Theme. In the examples above, the use of a parenthetical conjunctive adjunct (or 'partition', to use Taglicht's 1984: 22 term; see Chapter 4 for a definition) right after the topical Theme sets it apart from the rest of the T-Unit: if these examples were to be read aloud, the topical Themes would occur in their own separate tone units. By marking the topical Theme off from the rest of the T-Unit, the writer grants it a considerably greater amount of attention than it would have received if the CA had been placed in initial position. Compare, for example, (71) with its adapted version in (75), where *however* appears thematically: if one tries to read these two variants aloud, the degree of emphasis with which the rhematic 1 CA endows the topical Theme (as compared to its thematic 1 counterpart) becomes very clear. Thus, unlike thematic 1 uses, rhematic 1 CAs tend to draw attention not to the logical progression of discourse and the conjunctive relation themselves, but rather to another element in the T-Unit, viz. the topical Theme (see also Lenker 2011; 2014: 30 for similar observations). This emphatic discourse function of rhematic 1 CAs is even more striking when this position is combined with a very marked choice of topical Theme, such as the verb complement in (76).

(76) Gordon Brown's attack on his leader was premeditated. Attention has tended to concentrate on his repeated BBC interviews in which he violated all conventions, including that of common politeness, by complaining that the Prime Minister had not given him a post on Labour's National Executive Committee. <u>Much more serious</u>, **however**, was the Chancellor's article in The Telegraph, which, although arguably within the letter of Labour policy, amounted to an attack on Mr Blair's cherished Euro-constitution. (Mult-Ed-EN – The Telegraph)

In many cases, the emphasis placed on the topical Theme goes hand in hand with a desire to underline a topical shift occurring in discourse. Rhematic 1 CAs typically

occur when the writer is breaking the thematic continuity of the text, by moving on from the discussion of one topical element to that of another (see also Altenberg 2006: 19–21 on this). Such topical shifts may be circumstantial, as in (72): in this example, the typographical isolation of the marked topical Theme by means of a rhematic 1 CA gives the reader an explicit signal that we are no longer talking about 1945, but that the content of S2 rather concerns the year 2003. Likewise, in (74), the rhematic 1 use of *en revanche* highlights the topical shift from Peru to Bolivia. Interestingly, in such cases the thematic choice made by the writer can be said to be doubly marked: firstly, the function of topical Theme, which is normally reserved in priority to subjects, is occupied by an adjunct; in addition, the marked thematic choice is given yet more focus through being detached from the rest of the T-Unit by means of a parenthetical element (see also Virtanen 1992; Hasselgård 2010b; 2014b for more details on the discourse features of initial adjuncts). On the other hand, rhematic 1 CAs may also underline a topical shift from one (group of) participant(s) to another, as in (73), where the writer is making clear that s/he is shifting focus from Max Weber's electoral status to his wife's; or (71), where the researcher is shifting attention from the employment patterns of childless women to those of mothers. In this respect, it is useful to refer to Downing's (1991) view of topical Themes as 'frameworks'. According to Downing, topical Themes can be said to function as frameworks for the T-Unit, in that they set "the spatial, temporal or individual framework within which the main predication holds" (Chafe 1976: 50, quoted in Downing 1991: 128; see also Charolles & Vigier 2005; Hasselgård 2010b: 77; 2014b: 77 on the function of initial adverbials as discourse frameworks). By laying focus on such frameworks, rhematic 1 CAs indicate explicitly to the reader that the framework for interpreting the content in S2 is no longer the same as that which held in S1, thereby guiding the reader through some of the topical shifts taking place in discourse. A similar function of rhematic 1 CAs is also underlined by Lenker (2014: 32), who refers to rhematic 1 CAs as "delimitators for frame-setting adverbials".

In addition to highlighting topical shifts in discourse, however, rhematic 1 CAs may also be used to underline the thematic continuity of a text. For example, it is common for the fronted topical Themes to include one or more cohesive markers expressing reference and/or lexical cohesion (see Chapter 2 for a definition of the different subtypes of cohesive devices). In such cases, the topical Theme refers to entities which were mentioned earlier in the text, usually in the immediately preceding sentence(s) (see also Rossette 2009: 30 on the correlation between post-subject CAs and Themes conveying Given information). In (77), for instance, the topical Theme *that calculation* is a lexical resumption of the entire preceding sentence: the noun phrase 'encapsulates' the preceding sentence, i.e. it repackages an extended stretch of text in a shorter form (Flowerdew & Forest 2015: 5; see also Sinclair 2004: 83–85 for a definition of encapsulation). Likewise, in (78), the topical Theme *such a*

development summarises the information conveyed in the previous sentence. In (79), the cohesive reference conveyed in the topical Theme is slightly more far-reaching: the topical element preceding the rhematic 1 *however* stands for the eight studies discussed in the previous paragraph. In summary, in examples such as these, the rhematic 1 CA serves to reaffirm to the reader that the text is still dealing with the same referent as it was in the previous sentence or paragraph. Note that in many cases, the cohesive resumption used as topical Theme also sheds a new light on the previously-mentioned referent. In (80), for instance, *this leader-writer* is cohesively related to *Michael Fish* in the previous sentence. In addition to ensuring a cohesive link, such a formulation also provides a further specification of who Michael Fish is – or at least, how the writer perceives him. Similarly, while the topical Theme in (79) serves to refer to all the studies mentioned in the previous paragraph, it also adds the information that the rest of the T-Unit is going to talk about a quality that *none* of them displays.

(77)   The pro-Western governments were also hit hard, as tourism was immediately destroyed and local economies were ruined. This, extremists hoped, would foment discontent with governments hostile to the Islamists. That calculation, **however**, has often proved mistaken.

(Mult-Ed-EN – The Times)

(78)   If Ukraine can be drawn into the economic and political community of Europe without alarming Russia, this would be an immense achievement for European values and for the international standing of the EU. Such a development, **however**, will require the co-operation of Mr Putin.

(Mult-Ed-EN – The Financial Time)

(79)   Along similar lines, some scholars of social movements have stressed the importance of a formal, bureaucratic structure and clear division of labor in order to maximize movement outcomes and effective decision making (Gamson 1975; Zald and McCarthy 1987), while others have suggested a decentralized informal model to be the most effective, given the high adaptability and low vulnerability of a segmented structure to environmental changes (Lichterman 1995; Melucci 1996). Among them, Gerlach (2001) suggested a segmented, polycentric, and integrated structure-the so-called SPIN model-to be a powerful form of social movement organization, as it allows actors to prevent effective repression, penetrate into a variety of social niches through factionalism and schism, adapt to circumstances, and promote innovation. The few empirical studies of movements as social networks have not always generated consistent results: some have stressed the decentralized and multipolar nature of those networks (Sawer and Groves 1994), while others have shown movement networks to be highly centralized despite the informal nature of the ties linking movement actors to each other. (Lowe and Goyder 1983; Diani 2003)

> None of the contributions mentioned above, **however**, has proposed a formal account of the structure of political networks nor a systematic discussion of its properties (LOCRA-EN – Sociology).

(80)   If the Arctic air heading our way today and over the weekend produces a white Christmas, it will be one to remember, not another to forget. Michael Fish, betting £1,000 that it won't snow in London on Christmas Day, may revel in the role of national party pooper. This leader-writer, **though**, feels thoroughly restored by the prospect and is heading off in the best of seasonal spirits to a performance of Messiah. (Mult-Ed-EN – The Guardian)

To summarise, it is useful to refer to Hasselgård's (2004b: 77) statement that "[t]he beginning of a sentence is often seen as a strategically important position, where continuity as well as breaks in continuity can be marked". As shown in the previous paragraphs, rhematic 1 CAs are often used to highlight such signals of continuity and breaks in continuity conveyed by topical Themes. In this, rhematic 1 CAs may be viewed as reader-friendly devices: they typically help the readers navigate through the thematic progression of the text. It is also interesting to note that in the research articles, another common, equally reader-friendly function of rhematic 1 CAs is to underline discourse-structuring formulae referring to previous or following parts of the text, as in (81) and (82).

(81)   Critically, given the nature of these adaptive problems, the inputs that lead to sexual disgust should differ from those involved in the pathogen disgust system. Before we turn to an information processing model of sexual disgust, **though**, we briefly address the question, Why disgust?
(LOCRA-EN – Psychology)

(82)   In principle, it was enough to prevent fraud to make vendors return unsold tickets and sales receipts to the lottery promoter before the draw day. As we shall see in the next section, **however**, vendors could still devise strategies to cheat the promoter.   (LOCRA-EN – Sociology)

Another way in which rhematic 1 CAs may be said to be reader-friendly – which is closely related to their function as framework delimitators – is that they often isolate and clearly identify the entities that are being contrasted with the preceding discourse. In (83), for instance, the contrast lies between the Swedish as opposed to the English and French police officers; whereas in (84), what is being contrasted are the state of minds of migrants when they are in Italy as opposed to when they arrive in Sri Lanka. Likewise, in (85), the CA opposes the tactical and the political perspectives on the situation that is being described.

(83)   Une cinquantaine de voitures brûlées, une douzaine d'interpellations, quelques bâtiments vandalisés : le bilan de cinq soirées d'émeutes urbaines en Suède, depuis dimanche 19 mai, peut faire sourire dans les

> commissariats de Strasbourg ou de Brixton, en Grande-Bretagne, où la colère explose régulièrement et dans des proportions plus alarmantes. <u>Les Suédois</u>, **cependant**, prennent cette flambée de violences très au sérieux, et ils ont raison.   (Mult-Ed-Fr – Le Monde)

(84)   A main point of this essay is to understand why return migrants find it difficult to remain in their own community and soon re-migrate. After years of work in Italy, people want to return home, but not to live like impoverished fishermen. <u>Once in Sri Lanka</u>, **however**, their economic ambitions and their capacity for investment are discouraged, and they are mocked for "becoming Italian".   (LOCRA-EN – Anthropology)

(85)   D'un point de vue tactique, le gouvernement israélien aurait gagné à la traiter par l'indifférence. <u>D'un point de vue politique</u>, **en revanche**, M. Nétanyahou se croit sans doute contraint de donner des gages aux plus intransigeants de ses alliés.   (Mult-Ed-Fr – Le Monde)

(86)   Une cinquantaine de voitures brûlées, une douzaine d'interpellations, quelques bâtiments vandalisés : le bilan de cinq soirées d'émeutes urbaines en Suède, depuis dimanche 19 mai, peut faire sourire dans les commissariats de Strasbourg ou de Brixton, en Grande-Bretagne, où la colère explose régulièrement et dans des proportions plus alarmantes. **Cependant**, les Suédois prennent cette flambée de violences très au sérieux, et ils ont raison. (adapted version of Example (83))

The typographical isolation of these elements by means of the parenthetical CA helps the reader identify the main locus of the contrastive relation. If the same examples had been introduced by a thematic 1 CA, on the other hand, as in (86), the readers would have had to identify for themselves the elements being contrasted by the CA. In this respect, the function of rhematic 1 CAs is partly analogous to that performed by dislocated and cleft constructions (e.g. *it is…that/who; c'est… qui; Paul, he…*, etc.), which also serve to focus attention on a particular part of the message, usually by contrasting it to some other discourse entity (see Chapter 7 for more details). As already mentioned previously, such syntactic structures tend to be much more common in French than in English, where they usually sound very marked (see e.g. Rivelin-Constantin 1992; Carter-Thomas 2002; Lambrecht 2010). This cross-linguistic difference may in part explain why rhematic 1 CAs are markedly more frequent in English than in French: such uses of CAs may be viewed as one device used by English writers to compensate for a relatively rigid word order, which does not easily allow for constituents to be focused and/or contrasted through special syntactic structures such as clefting or dislocation. French, on the other hand, which more easily tolerates such syntactic constructions, may have more limited use for rhematic 1 CAs. In a context such as the one in Example (87), for instance, a French writer would probably have used an emphatic cleft structure

to obtain a similar effect to that produced by the rhematic 1 CA, such as: *Néanmoins, ce sont les enfants qui ont raison*. That being said, it is also worth noting that the French subcorpus displays a noticeable frequency of examples combining the use of a rhematic 1 CA with a dislocated syntactic structure, as in (88) and (89).

(87)   Today's Guardian/ICM poll […] suggested that British Muslims were surprisingly optimistic about the position of themselves and their families, with 44% saying that they expected life to get better as against 33% who said it would get worse. Paradoxically, the Guardian conference that brought together the brightest of a new generation of young Muslims nominated by a wide cross-section of Islamic groups, were less optimistic about the future than those in the poll. They were gloomy over the prospects of "the war on terror", while the idea that British Muslims might play an equal, valued role in British society by 2010 was dismissed as "hopelessly optimistic". Remember this was a group of university-educated professionals, who had already advanced to positions their grandfathers and fathers never expected, though they worked unremittingly hard to ensure their children got there. The children, **however**, are right. They are the lucky "tadpoles" who succeeded, but most of their contemporaries will be suffering the poverty, discrimination and deprivation that the Muslim community has faced since serious numbers began migrating here four decades ago.

(Mult-Ed-EN – The Guardian)

(88)   Ces initiatives sont-elles individuelles ou collectives ? Pour l'heure, la réponse n'est pas claire. <u>Ce que l'on sait</u> **en revanche**, c'est que l'Etat islamique est une mouvance très organisée et hiérarchisée, qui ne laisse rien au hasard.   (Mult-Ed-FR – Libération)

(89)   De l'indispensable réduction des dépenses de l'État, des collectivités et des administrations, le programme Hollande ne dit mot, ou presque. <u>C'est</u> **pourtant** <u>par là qu'</u>il aurait fallu commencer avant d'élaborer un tel matraquage fiscal.   (Mult-Ed-FR – Le Figaro)

In his article, Altenberg (2006: 20) noted that rhematic 1 CAs frequently mark a double contrast: one between the fronted topical Theme and a preceding element in discourse; and a second one between what follows the CA in S2 and another part of S1. This pattern is also observed in the corpus analysed here. In (84) above, for instance, the contrastive CA opposes not only the temporal adjuncts (viz. when in Italy vs once in Sri Lanka), but also the ambitions of the migrants in each of these settings. Likewise, in (85), the CA opposes both the viewpoint adjuncts (from a tactical vs from a political point of view), and the Israeli government's most suitable attitudes from each of these two perspectives. In such cases, the rhematic 1 marker thus acts as a boundary between the two terms of the contrast.

Interestingly, as the reader may remember, this kind of binary contrast was one of the characteristics of the 'opposition' subtype of contrastive meaning, which typically opposes two pairs of constituents (as in: *John is tall <u>but</u> Mary is short*; see Chapter 2 for more details). Therefore, there seems to be some sort of correlation between the semantic relationship of opposition and the rhematic 1 position. However, this does not mean that CAs expressing other subtypes of contrast never occur in rhematic 1 position. Examples (90) and (91) below, for instance, illustrate rhematic 1 uses of concessive CAs. In fact, a cursory glance at the corpus data seems to suggest that, whereas rhematic 1 CAs which highlight a topical shift in discourse – as in Examples (83) to (85) above – frequently express opposition, rhematic 1 CAs that occur in contexts of thematic continuity – as in Examples (77) to (80) above – tend to express relationships of concession. More systematic semantic analyses are nevertheless required in order to corroborate this observation. Also note that in some examples, lexical factors also seem to be at play in determining the position of the marker: in (80) and (81), for instance, *though* could not have occurred sentence-initially, as this marker can only be used at the beginning of the sentence when it functions as a subordinator. The role played by lexis on CA placement is further discussed in Section 8.5.

(90)   In Gellis (2012), adherence was high and significant improvements over time were observed in sleep quality, insomnia severity, and various sleep parameters. <u>That investigation</u>, **however**, tested the intervention with a noncontrolled design using a small sample.   (LOCRA-EN – Psychology)

(91)   Sa stratégie, elle est sans fioritures : les Français, croit-il, sont avec le gouvernement parce qu'ils ne veulent plus des régimes spéciaux. Et sont donc contre la grève. <u>Le calcul</u>, **pourtant**, est dangereux.

(Mult-Ed-FR – Libération)

Finally, it is also important to note that in some cases, the rhematic 1 position was encouraged by the presence of another conjunctive marker at the beginning of the T-Unit. As the thematic 1 position was already occupied, the writer sometimes had to place the conjunctive adjunct later in the sentence. In some cases, the initial position was filled by another conjunctive adjunct of contrast – such as *however* in (92) – or, more commonly, by a coordinator – as in (93). In yet other examples, the element blocking the initial position was a concessive subclause, usually functioning as the first segment in the contrastive relation, with the CA reinforcing the contrastive link expressed by the subordinator, as in (94). Depending on the examples, using the conjunctive adjunct right before or after the other conjunctive marker would have been either stylistically (more) infelicitous – as in (93) – or simply grammatically incorrect – as in (92).

(92)   This reflexive process, or "journey inward" as coined by Conway and Clark (2003), entailed the deconstructing and restructuring of their own identities, which gave the novice teachers a great challenge. <u>However</u>, it **nonetheless** helped them to find clearer goals and take more intrinsically motivated. (teaching) actions and brought new understandings of themselves in a higher degree of self-awareness   (LOCRA-EN – Education)

(93)   To be fair, this Labour Government has made fewer defence cuts than its Tory predecessor. <u>But</u> those cuts **still** mean that the defence budget has halved in real terms over the past decade, even though September 11 introduced us to new and present dangers.   (Mult-Ed-EN – The Telegraph)

(94)   S'il est vrai que les révoltés modernistes se sont révélés à la fois minoritaires et divisés (sans ces divisions, ils seraient majoritaires), **en revanche**, il est essentiel de remarquer que, pour la première fois, les islamistes utilisent le mot "démocratie" de manière positive et comme un idéal avec lequel ils prétendent concilier l'islam.   (Mult-Ed-FR – Le Nouvel Observateur)

In summary, a qualitative examination of the CAs occurring in rhematic 1 position uncovered three main, closely intertwined discourse effects performed by such items: (i) laying focus on the detached topical Theme; (ii) emphasising thematic transitions in discourse, viz. most frequently a topical shift but also sometimes a certain amount of thematic continuity in the text; and (iii) clearly identifying the entity being contrasted by means of the contrastive CA. Depending on the examples, these effects were at play either simultaneously, or separately – with the first of these three functions being the most constant across the examples in the corpus.

8.4.1.3   *Rhematic 2 conjunctive adjuncts*

Similarly to rhematic 1 CAs, rhematic 2 conjunctive adjuncts – exemplified in (95) to (100) – tend to be associated with a range of rhetorical effects pertaining to emphasis, thematic and information structure. Firstly, in line with Lenker's (2011; 2014) earlier findings, most rhematic 2 CAs in the English subcorpus were found to draw increased attention not to the Theme, but to the Rheme of the T-Unit. Through being used parenthetically within the predicate, many rhematic 2 CAs qualify as one of the three devices identified by Taglicht (1984: 20) to create a so-called 'marked Rheme', viz. a rhematic element that is syntactically and/or typographically detached from the elements with which it would be contiguous under usual circumstances (see also Chapter 4 for more details on marked Rhemes). Because rhematic 2 CAs interrupt the usual 'verb – complement' sequence of the message, the dependency structure of the T-Unit is momentarily left incomplete, while the reader is kept waiting for the syntactically obligatory complement, or even for the end of the verb phrase. As argued by Taglicht (1984: 25), this element of delay typically grants the Rheme more focus than it would have received in the equivalent unmarked sequence. Compare,

for instance, (95) with its adapted version in (101); or (96) with (102). As the Rheme in these adapted examples is fully integrated into the rest of the T-Unit, it receives far less focus than in the original extracts.

A close examination of rhematic 2 CAs in French, on the other hand, revealed that, as opposed to their English counterparts, these markers are most often typographically integrated into the rest of the T-Unit. This is clearly visible from the examples below: out of the four instances of rhematic 2 CAs in French, the only CA used parenthetically is *cependant* in (98). As a result, the focusing effect associated with rhematic 2 CAs appears to be much less salient in the French than in the English subcorpus. This greater syntactic integration of rhematic 2 CAs in French as compared to English may be interpreted as further evidence of the fact that, as was already clear from the comparison of frequencies of use of each position in the two languages, this syntactic slot is central in the French linguistic system, whereas it is more marginal – or indeed marked – in English. In other words, placing a CA of contrast within the predicate appears to be so natural in French that such markers do not even require to be typographically isolated from the dependency structure of the T-Unit.

(95) As Moses and Ryan (2006) discussed, the evidence for a distinction between conjunctive (intrinsic) and relational (extrinsic) binding is most compelling in the field of long-term episodic memory. The present article argues, **however**, that the distinction between intrinsic and extrinsic binding is not only relevant for episodic long-term memory but is a principle that applies also to short-term working memory.   (LOCRA-EN – Psychology)

(96) While the union leaders denounced the independent review conducted by Sir George Bain they were quietly confident that his report would provide the vehicle for extorting cash from the employers. The Bain findings were, **however**, carefully balanced.   (Mult-Ed-EN – The Times)

(97) C'est surtout dans les communes rurales que l'on a tendance à penser que la société est injuste. Si l'on ajoute la catégorie socioprofessionnelle, on constate que ce jugement négatif sur la société dans son ensemble est en premier lieu le fait des catégories populaires (ouvriers ou employés, et davantage encore pour les ouvriers non qualifiés) résidant dans ces communes rurales ou dans des villes moyennes ou petites (à l'inverse des cadres supérieurs qui dans ces zones continuent de trouver la situation plus juste). Ce sentiment d'injustice est **en revanche** beaucoup moins prégnant parmi les ouvriers ou employés des banlieues ou des grandes villes.
(LOCRA-FR – Sociology)

(98) Avec ses combats "à l'ancienne", filmés à l'aide d'un nombre réduit de caméras, Skyfall souligne, à sa manière, la désuétude de cet émissaire d'un autre temps, ressortant du garage sa sempiternelle Aston Martin, qu'il

conduisait déjà il y a un demi-siècle. Et incarnant, sans désemparer, la résistance de l'homme blanc face au déclin de l'Occident. Cette excentricité toute britannique ne doit pas masquer, **cependant**, l'étonnante plasticité de la saga, prompte à s'adapter aux mutations de notre temps.

(Mult-Ed-FR – Le Monde)

(99) La cohérence que nous recherchions en début de notre enquête n'est donc pas là où nous la supposions : au cœur de l'un ou l'autre de ces stratagèmes de communication. Elle se niche **au contraire** dans leur juxtaposition.

(LOCRA-FR – Sociology)

(100) Fernand refuse de parler de son enfance en entretien. Tout juste saura-t-on qu'il est né en 1981, qu'il a été élevé à Alger par sa grand-mère et qu'il a développé à ses côtés. Fernand est **en revanche** plus prolixe lorsqu'il s'agit d'évoquer sa profession. (LOCRA-FR – Political science)

(101) As Moses and Ryan (2006) discussed, the evidence for a distinction between conjunctive (intrinsic) and relational (extrinsic) binding is most compelling in the field of long-term episodic memory. **However**, the present article argues that the distinction between intrinsic and extrinsic binding is not only relevant for episodic long-term memory but is a principle that applies also to short-term working memory. (adapted version of Example (95))

(102) While the union leaders denounced the independent review conducted by Sir George Bain they were quietly confident that his report would provide the vehicle for extorting cash from the employers. **However**, the Bain findings were carefully balanced. (adapted version of Example (96))

While the focusing function of rhematic 2 CAs is mostly associated with the English markers, the corpus analysis also sheds light on two other rhetorical functions of conjunctive adjuncts occurring within the predicate, which were found to be more widely shared by the two languages. The first of these functions pertains to the notions of Given and New information. A large portion of the rhematic 2 CAs in the corpus were found to occur at the boundary between the information that is already known by the reader (i.e. Given information) and the genuinely new informational contribution of the message (i.e. New information; see also Lenker 2011; 2014 for similar findings). This is particularly clear in Example (96) above, for instance, where the first part of the T-Unit, preceding the CA, refers to an entity which was already mentioned in S1, viz. *the Bain findings*.[10] In the second part of the T-Unit, on the other hand, the writer provides some New information about

---

10. Following Lenker (2011: 17), copular verbs, auxiliaries and other informationally empty words are not considered to contribute new information to the T-Unit. Rather, the lexical verb and the complements are the major repositories of new information.

these Bain findings, viz. s/he states that they were *carefully balanced.* The same phe-
nomenon is visible in (98), where the main experiential information preceding the
CA, viz. *cette excentricité toute britannique*, corresponds to the Given information:
this noun phrase actually summarises – or, once again, encapsulates – the entire
content of the two previous T-Units, which described one particular aspect of the
latest *James Bond* movie. The informational status of this sequence is made very
clear by the demonstrative determiner *cette* that introduces it. After *cependant,* on
the other hand, the writer provides us with information that complements this
known content, telling us that in addition to providing a somewhat old-fashioned
picture of Great Britain, the movie was also able to adapt to the modern world.
Similarly in (97), the noun phrase preceding the CA, viz. *ce sentiment d'injustice*,
conveys Given information: it refers to an experiential entity which was not only
discussed in the two preceding T-Units provided in the example, but is also the
focus of the whole section of the article from which the example is extracted (enti-
tled: *Le sentiment de justice sociale*). After the CA, on the other hand, the reader
learns something new about the feeling of injustice that has been at the core of the
past few paragraphs. It is also worth noting that, although the Given information
is often presented at the left of the CA, with the New information following the
conjunctive marker, the corpus data also contained some instances presenting a
marked information focus, with the New information appearing at the left of the
CA, as in (103).

(103)   Des décennies de conflit israélo-palestinien le prouvent : de la première
         Intifada en 1987 à la prise de contrôle du Hamas vingt ans plus tard, il est
         sot ou vain d'occulter le sort de ce territoire. Nombreux sont **pourtant** ceux
         qui ont préféré nier les leçons de l'Histoire.   (Mult-Ed-FR – Le Monde)

In summary, the detailed analysis of English and French rhematic 2 CAs corro-
bates Lenker's (2011: 17; 2014: 113) earlier findings, by showing that in addition
to performing their basic linking function as markers of contrast, these items
also tend to act as "partitioners of given and new information", showing explic-
itly to the reader where the boundary between Given and New information lies
in the T-Unit. In this respect, the results provide empirical support in favour
of Halliday's (1985: 81) early, largely intuition-based observation that, when
they occur non-thematically, conjunctive adjuncts "tend to occur at points in
the clause that are significant for textual organization, which means at some
boundary or other". As explicit signals of a boundary between Given and New
information, and in the same way as rhematic 1 markers, rhematic 2 CAs may
also be viewed as reader-friendly devices: they provide the reader with a clear
indication on where to find the information that is going to make the message
move forward.

A third major, recurrent function of rhematic 2 CAs is that they typically isolate and/or identify the main locus of contrast, viz. they point to the main experiential content that is being contrasted with preceding discourse, separating it from the rest of the T-Unit. In this, the corpus observations are therefore partly consistent with Greenbaum's (1969: 195) study on the placement of attitudinal disjuncts in English, in which the author found that non-initial attitudinal disjuncts tend to be placed "as near as possible to the major information point [to help] focus attention on it" (see also De Cesare 2013 for a similar observation on Italian). In (96) above, for instance, it is the *carefully balanced* character of the Bain findings – i.e. the information that follows the rhematic 2 CA – that is being contrasted with the expectation, expressed in S1, that these findings would constitute a perfect opportunity for union leaders to extort cash from employers. What precedes the CA, on the other hand (viz. *The Bain findings*), simply constitutes the information that is common to the two segments related by *however* (i.e. the Given content), and is in no way contrasted to any part of the preceding T-Unit. Likewise, in (98), the information that follows the CA (viz. the astonishing plasticity of the *James Bond* saga and its ability to adapt to present times) is the element that the writer is opposing to the rather old-fashioned attributes of the movie depicted in the previous sentences. What precedes the CA (viz. *cette excentricité toute britannique*) is not contrasted to the content expressed in S1: rather, as already pointed out above, it summarises S1. In (100), too, the two elements which stand in contrast with what precedes (viz. (i) *plus prolixe* and (ii) *lorsqu'il s'agit d'évoquer sa profession*, which contrast with (i) *refuse de parler* and (ii) *de son enfance*, respectively) are placed after the CA. In other words, rhematic 2 CAs are placed close to the informational segment over which they have scope. This property of rhematic 2 CAs is particularly clear with the markers occurring between a projection verb and the following *that*-clause, as in (104) below. In this example, the element in S2 that truly contrasts with the preceding discourse is the possibility of an evolution of attitudes towards group work and tutoring in higher educational settings. In fact, from a purely discourse-organisational perspective, the adapted example in (105) would have been roughly equivalent to the original wording: the contrastive relation in (105) remains virtually identical to that in (104), which goes to show that the CA really mainly has scope over the segment that follows it.

(104)   Tout les amène logiquement à considérer que la qualité des relations au sein des groupes est difficile à observer (manque de temps, manque de catégories d'analyse) et moins importante que les connaissances techniques acquises et l'efficacité productive. Tous ces éléments se répercutent sur leur perception du rôle attendu du tuteur. L'étape finale de la recherche-action démontre **toutefois** qu'une évolution est possible.   (LOCRA-FR – Education)

(105)   Tout les amène logiquement à considérer que la qualité des relations au sein des groupes est difficile à observer (manque de temps, manque de catégories d'analyse) et moins importante que les connaissances techniques acquises et l'efficacité productive. Tous ces éléments se répercutent sur leur perception du rôle attendu du tuteur. **Toutefois**, une évolution est possible.

<div align="right">(adapted version of Example (104))</div>

The fact that rhematic 2 CAs tend to be placed close to the linguistic segment(s) over which they have scope, thus isolating the main informational focus of the contrastive relation, constitutes another way in which rhematic 2 uses of conjunctive adjuncts may be viewed as reader-friendly devices. If these CAs had been used thematically, as in the adapted version of Example (96) provided in (102) above, the readers themselves would have had to identify the specific elements being contrasted by the markers within the whole T-Unit. Thus, in line with Greenbaum's (1969: 195) observation that "[t]he placement of the disjunct in written English can be of importance in providing a clue to the reading of the clause", rhematic 2 CAs help readers identify the scope of the contrastive relation being conveyed by the conjunctive marker.

   In summary, a close examination of rhematic 2 CAs in the corpus data sheds light on three main – and once again closely interrelated – discourse factors that may trigger such a positional choice, viz. a wish to (i) lay focus on the Rheme of the T-Unit; (ii) set an explicit boundary between Given and New information; and/ or (iii) isolate the main locus of contrast in the T-Unit. In some of the examples analysed in the corpus, these three rhetorical effects were at play simultaneously, but in other examples, only one or two of them were observed. Finally, for the sake of completeness, it is also important to stress that, as was the case with rhematic 1 CAs, the choice of the rhematic 2 position is sometimes (also) induced by the presence of another conjunctive marker (viz. coordinator, conjunctive adjunct or concessive subclause) in initial position, as in (106) and (107). In addition, in some cases, the choice of the rhematic 2 position seemed to be motivated less by rhetorical than lexical factors. Most instances of the French marker *plutôt,* for example, would have sounded unacceptable, or at least fairly unnatural in thematic 1, or even in rhematic 1 positions (see Examples (108) and (109)).

(106)   La production de Rosemary Sayigh et de Michael Johnson n'a pas été « suspendue » avec autant de radicalité, <u>mais</u> elle demeure **néanmoins** intéressante car l'un et l'autre anthropologues ont écrit sur le problème de l'analyse des données dans un contexte radicalisé par la violence politique.

<div align="right">(LOCRA-FR – Anthropology)</div>

(107)   And <u>even if the attempt to hold the Games ends in failure</u>, then the money put into the enterprise will **still** have done more good than many of the grants that have been made from the lottery.   (Mult-Ed-EN – The Times)

(108)   Valéry Giscard d'Estaing, François Bayrou, Robert Badinter, se gardent bien de mettre en cause l'appartenance de la Turquie à l'Islam, sur lequel l'Empire ottoman a d'ailleurs régné plusieurs siècles : leurs électeurs, en revanche, ont peur de l'entrée en Europe d'un pays musulman de bientôt 80 millions d'habitants. Ils devraient **plutôt** se réjouir qu'un grand Etat musulman souhaite si ardemment intégrer l'Europe chrétienne de 450 millions d'habitants et soit prêt à en adopter les valeurs.   (Mult-Ed-FR – Libération)

(109)   Cela ne signifie nullement d'ailleurs que l'eau ruisselait sur les parois de la rotonde sud de Baume Peinte. Une autre expérimentation conclut **plutôt** en faveur de supports secs.   (LOCRA-FR – Anthropology)

As already mentioned above, the influence of lexis on the placement patterns of English and French CAs of contrast is discussed in more detail in Section 8.5.

8.4.1.4   *A short word on thematic 2 and rhematic 3 conjunctive adjuncts*

As the thematic 2 and the rhematic 3 positions only represent a small fraction of the CAs in the corpus (viz. 7% in English, and just over 2% in French), and because they play a more limited role in distinguishing both (i) languages and (ii) registers than the other three positions (see Section 8.3), the discourse effects that they produce are only discussed very briefly here. The objective is mainly to make clear that each of the five positions identified in this chapter is partly specific in terms of the discourse functions that it fulfils, and that the thematic 1 position is truly the only position performing a purely conjunctive function.

Thematic 2 conjunctive adjuncts, which occur right after an interpersonal Theme, are exemplified in (110) and (111).

(110)   A '10.50 levy comes into force next year on every home insurer. It means we all pay for flood damage since potentially any one of us could also find ourselves in possession of "a stigmatised asset". <u>Rightly</u>, **however**, the coverage is not available for new builds on flood plains and in B and H of the council tax.   (Mult-Ed-EN – The Observer)

(111)   Cependant, parce que peu d'information est disponible dans les études sur le nombre de mois passés en institution et l'âge de l'enfant à l'arrivée en institution, il est difficile de bien saisir la signification de ce résultat. <u>Il demeure clair</u>, **néanmoins**, <u>que</u> la possibilité de développer un attachement sécurisant est mince.   (LOCRA-FR – Psychology)

Such markers produce a roughly similar effect to that of rhematic 1 CAs, viz. they lay emphasis on the element that precedes them. The use of a parenthetical CA right after an initial interpersonal element sets it apart from the rest of the T-Unit, granting it its own separate tone – or rather, typographical – unit. This has the effect of bringing the personal stance of the writer to the fore, which usually takes

precedence over the conjunctive link. In (110), for instance, the fronting and the typographical isolation of the stance adverb *rightly* draws attention to the opinion of the writer towards the content of the clause. This in turn orients the reader's entire assessment of what follows in the way intended by the writer: before even reading the content of the T-Unit, the reader is prompted to interpret it as being 'right'. As is clear from the adaptation of the example in (112) below, such an effect would not have been obtained had *however* been used initially, especially as the stance adverb would then have had to be placed within the predicate, thus receiving much less focus than in the original example. In a similar vein, the typographical isolation of the impersonal clause *il demeure clair* in (111) also highlights the high degree of certainty of the writer towards their statement. Again, this stance meaning would not have received as much emphasis if the CA had been used initially, as in (113).

(112) A '10.50 levy comes into force next year on every home insurer. It means we all pay for flood damage since potentially any one of us could also find ourselves in possession of "a stigmatised asset". **However**, the coverage is <u>rightly</u> not available for new builds on flood plains and in B and H of the council tax.   (adapted version of Example (110))

(113) Cependant, parce que peu d'information est disponible dans les études sur le nombre de mois passés en institution et l'âge de l'enfant à l'arrivée en institution, il est difficile de bien saisir la signification de ce résultat. **Néanmoins**, il demeure clair que la possibilité de développer un attachement sécurisant est mince.   (adapted version of Example (111))

Interestingly, by virtue of the fact that they can be used to bring focus to surrounding elements in the T-Unit, thematic 2 CAs are stylistically closer to rhematic markers than they are to thematic 1 CAs, which are fully restricted to their primary linking function. Therefore, from a stylistic perspective, one may wonder how justified it is to group thematic 2 markers in the same category as the CAs occurring at the very beginning of the T-Unit. This theoretical issue would deserve further investigation in future research.

A detailed investigation of CAs occurring in rhematic 3 position revealed that, as was the case for the rhematic 2 and, especially, the rhematic 1 categories, this positional choice is commonly motivated by factors having to do with focus and thematic progression. In many examples from the corpus, the use of a CA in rhematic 3 position seems to be motivated by a (probably unconscious) will of the writer to give precedence to the topical over the textual sequencing of discourse. In Examples (114) and (115), for instance, the topical Themes in the T-Units hosting the conjunctive adjunct both ensure thematic continuity with the preceding discourse. In the first example, the phrase *the philosophical confusions of Ukip*

summarises, or 'encapsulates' (cf. above) the whole content expressed in the previous paragraph and is then chosen as the point of departure of the next paragraph. Likewise, in (115) – which also illustrates the non-final version of the rhematic 3 position; see Section 8.2 for more on this distinction – *cette percée* paraphrases the content expressed in the preceding sentences. It refers to the same referent as previous phrases, among which *cette tentative* or *une petite révolution*, hence contributing to the lexical cohesion of the passage (see Chapter 2 for a definition of lexical cohesion).

(114)   It is revealing, incidentally, that Ukip is willing to sacrifice its best chance of a referendum, and likely install a Labour government, for the sake of its own electoral success. It shows how much the party has changed since 2009, when its then leader, Lord Pearson, told Mr Cameron that he would disband Ukip altogether if the Tories offered a referendum. Now Ukip's chief motivation appears to be ensuring its own survival. As we report today, research from Business for Britain shows that our MEPs are near powerless in Europe; given our voice there is smaller than ever, it seems all the more baffling that Ukip has chosen to put party before country.

The philosophical confusions of Ukip do not change the facts, **however**: the party is still a frustrating obstacle to Mr Cameron, and to his efforts to reform the EU (Mult-Ed-EN – The Telegraph).

(115)   Même balbutiante, cette tentative baroque d'élection interne est une première dans l'histoire d'une droite jusqu'ici tout entière vouée au culte du chef. Les militants auront le dernier mot et c'est en soi une petite révolution dans un parti où le chef s'est toujours imposé sans que nul n'ait pu en disposer. Il fallut un désastre électoral pour que la droite se résigne à mettre elle aussi sa tête aux voix. Cette percée resterait vaine **pourtant** si elle n'allait de pair avec une clarification devenue indispensable après la fin de campagne extrême de Sarkozy.   (Mult-Ed-FR – Libération)

In contexts such as these, the decision to use a CA at the end of the T-Unit may perhaps be ascribed to the writer's wish to set the topical element as the only point of departure of the message, making sure that it does not have to share thematic status with any other unit. As pointed out by Hasselgård (2010b: 74), "[t]hematic status implies some degree of focus". Therefore, if the CAs in the examples above had been placed in thematic 1 position, as in (116), some of the thematic focus would have been diverted away from the topical Theme to the conjunctive link. When the CAs are placed in rhematic 3 position, on the other hand, the contrastive link is downplayed so as to direct the reader's full attention to the topical element as s/he moves on from one sentence to the next, hence highlighting thematic continuity in the text.

(116)   It is revealing, incidentally, that Ukip is willing to sacrifice its best chance of a referendum, and likely install a Labour government, for the sake of its own electoral success. It shows how much the party has changed since 2009, when its then leader, Lord Pearson, told Mr Cameron that he would disband Ukip altogether if the Tories offered a referendum. Now Ukip's chief motivation appears to be ensuring its own survival. As we report today, research from Business for Britain shows that our MEPs are near powerless in Europe; given our voice there is smaller than ever, it seems all the more baffling that Ukip has chosen to put party before country.

However, the philosophical confusions of Ukip do not change the facts: the party is still a frustrating obstacle to Mr Cameron, and to his efforts to reform the EU.   (adapted version of Example (114))

In addition to highlighting the thematic *continuity* of a given text, rhematic 3 CAs may also be used to emphasise breaks in thematic continuity. In (117), for instance, the topic of the passage shifts from *Mr Bush* in the first two sentences, to *Europe's leaders* in the third one. Again, by being established as the single point of departure of the message, the topical Theme in this last T-Unit receives the totality of the thematic focus, which is even further reinforced by the presence of the parenthetical adverb *too* that directly follows it. In a similar vein, Example (118), which reproduces the main segments of part of an editorial from *Le Figaro* discussing the unemployment rate in France, is particularly representative of the type of context in which rhematic 3 CAs are found. The first part of the example presents one aspect of the issue at hand, viz. that the current situation of the labour market is unsatisfying, for a number of reasons. The T-Unit that contains the rhematic 3 CA then operates a topical shift in discourse, moving on from what has been going wrong to more encouraging aspects of the employment situation. Again, the contrastive link expressed by *néanmoins* is relegated to the end of the sequence, so as to grant the entire thematic focus to the noun phrase expressing this thematic shift, viz. *des signes encourageants*. As is often the case, the rhematic 3 CA is part of a very short segment, whose content is then further developed in the sentences that follow. In Example (118), the sentence containing the CA thus functions as a 'topical pivot', as it were, with the content expressed by the topical Theme (viz. *des signes encourageants*) being discussed in greater detail in the directly following discourse.

(117)   Mr Bush stands a chance of success only if allies help, and only if he reaches out to allies. He can and should show his multilateral side by pushing hard for a deal in the World Trade Organisation. Europe's leaders, too, need to face up to reality, **however**.

(118)   Mais la situation de l'emploi est-elle pour autant satisfaisante ?
         Malheureusement non, et ce, pour au moins trois raisons. D'abord parce
         que notre taux de chômage demeure nettement supérieur à la moyenne
         européenne. […] Ensuite parce que notre pays souffre toujours du fait que
         la part de la population qui travaille est plus faible qu'ailleurs […] Enfin
         parce que le nombre de personnes sous-employées mais qui voudraient
         travailler davantage progresse depuis 2006, c'est-à-dire depuis que la courbe
         du chômage baisse. […] On le voit bien, c'est à ces spécificités françaises
         qu'il convient de s'attaquer. L'objectif fixé par le président de la République
         pour son quinquennat – 5 % de chômeurs et 70 % de Français au travail
         – montre le chemin qui reste à parcourir. L'objectif fixé par le président
         de la République pour son quinquennat – 5 % de chômeurs et 70 % de
         Français au travail – montre le chemin qui reste à parcourir. Des signes
         encourageants apparaissent **néanmoins.** Le taux d'emploi, c'est-à-dire la
         part des Français qui travaille, commence à remonter. Est-ce le début d'un
         cercle vertueux? Autre signe, depuis la rentrée, les partenaires sociaux
         discutent d'une réforme des conditions d'embauche et de licenciement, de
         l'évolution des contrats de travail, bref s'efforcent de combler le fossé entre
         ceux qui occupent un emploi à durée indéterminé et les autres. Ils le font
         dans un esprit constructif et le gouvernement fait bien de leur laisser un peu
         de temps pour aboutir. C'est l'organisation même du travail qui est en train
         de bouger dans notre pays. Et cela, en soi, vaut tous les bons chiffres du
         chômage.   (Mult-Ed-FR – Le Figaro)

In summary, a large portion of the rhematic 3 positional choices appeared to be
motivated by a wish to emphasise the thematic (dis)continuity of a given passage.

To summarise the observations made in Section 8.4.1, a more detailed, quali-
tative scrutiny of the patterns of English and French conjunctive adjunct place-
ment in the corpus revealed that this phenomenon is not simply a matter of
syntax. Rather, CA placement was shown to be associated with a range of rhetori-
cal functions at discourse level, having to do with focus, emphasis, information
and thematic structure. While CAs in thematic 1 position were shown to perform
a strictly linking function, all the other positions appeared to also endorse further
functions in addition to their basic connecting role, by drawing attention to other
parts or aspects of the message, such as the topical Theme (rhematic 1 CAs), a
boundary between Given and New information (rhematic 2 CAs), or the stance of
the writer (thematic 2). In other words, as pointed out by Lenker (2014: 30), when
CAs are not used initially, they "[do] not draw attention to the (unusually placed)
adverbial itself, but to other parts of the sentence". It is perhaps in this sense that the
thematic 1 position may be characterised as the 'unmarked' or the 'default' posi-
tion for conjunctive adjuncts, as is usually the case in the literature. In Section 8.3,
it was shown that in terms of frequencies, it was somewhat oversimplistic to

describe the thematic 1 position as the 'unmarked' or the 'normal' position for conjunctive adjuncts of contrast: in the editorials, for instance, this position was not even the most frequent one, and accounted for less than 40% of the CA tokens. Functionally, on the other hand, the thematic 1 position may in fact be qualified of unmarked, as it is the position that is used unless other elements justify selecting another slot. Choices pertaining to conjunctive adjunct placement thus constitute a characteristic example of Halliday's (1971: 174) so-called 'good reason principle': the thematic 1 position may be said to be unmarked in the sense that it "is the one which is chosen to express the meaning in question unless there is good reason to choose otherwise" (ibid.).

In addition, the observations made in this section also drew attention to the necessity of reconsidering the traditional view of conjunctive adjuncts as purely linking devices. Instead, as argued by Lenker (2011: 9), CAs also constitute "a particularly apt means for fulfilling further discourse functions in addition to their basic connecting function". As explained at the beginning of this section, grammars and reference books typically mention the potential of conjunctive adjuncts for syntactic mobility, while providing very few – if any – indications on the reasons that should motivate users to choose one positional option over the other in a particular context. Rather, in both the reference and the pedagogical literature, the focus tends to be mainly (if not solely) on the linking functions of such items. Yet, the analyses presented in this section have demonstrated the great rhetorical potential that is inherent in CA placement choices. In the future, it would be worth drawing more attention to such aspects of CA use, as they may constitute an invaluable resource for native writers and learners alike. This echoes a plea previously made by Hawes and Thomas (2012), who highlighted that issues pertaining not only to thematic structure, but also to the rhetorical effects achieved by thematic choices, are an essential component of successful, idiomatic writing, which tends to be largely neglected in teaching.

Furthermore, the fact that the different positions were demonstrated to perform their own specific (set of) rhetorical effects provided further justification in favour of the decision to make distinctions within the Rheme: while in Section 8.3, it was demonstrated that viewing the Rheme as one single entity would have led me to overlook significant cross-linguistic differences in CA placement, the qualitative discussion presented in this section also showed that the various syntactic slots available within the Rheme are not fully equivalent in terms of the discourse effects that they produce – a fact which would also have remained unnoticed had the Rheme been viewed as a single entity. Finally, it is important to specify that, although this section has mostly focused on the rhetorical and discourse effects motivating the use of one position over the others, I am very well-aware that these are not the only relevant elements to understand the phenomenon. For example,

factors of a more strictly linguistic nature, such as the length of the subject, or the complexity of the verb phrase, may also play a part in determining placement choices (see e.g. Osborne 2008 on the influence of the linguistic environment on adverb placement in learner and native English; Hasselgård 2010b for similar considerations regarding adjunct placement). Such analyses are beyond the scope of the present study, but certainly constitute a very promising avenue for future research.

8.4.2   Discourse effects of CA placement and cross-register differences

In the previous section, it was demonstrated that the different syntactic positions available for conjunctive adjuncts are not equivalent from a rhetorical point of view, but that they are each associated with their own range of discourse effects. In this section, I argue that the various rhetorical effects produced by CA placement partly explain the striking difference in ratio of thematic and rhematic markers found between the academic and the editorial registers in each language. As a reminder, the corpus analysis revealed that in both English and French, the majority of conjunctive adjuncts of contrast in the academic subcorpus occurred in thematic 1 position – as in (119) and (120), whereas the editorials displayed a much stronger preference for the rhematic positions than academic prose. Depending on the language, the preferred rhematic position differed: whereas the rhematic 1 position – exemplified in (121) – prevailed in the English editorials, in the French editorials the majority of CAs of contrast were found in rhematic 2 position – as in (122).

(119)   Following food-exposure, restrained and normal eaters displayed the same level of hedonic response at short ISI. **However**, whereas the hedonic response of normal eaters was down-regulated at the longer ISI, that of restrained eaters was not.   (LOCRA-EN – Psychology)

(120)   Saporta et Verstraete (2000) avancent que l'enseignement de l'entrepreneuriat peut modeler la cognition de l'étudiant en favorisant la combinaison de trois dimensions irréductibles et indissociables : réflexion, réflexivité et apprentissage. **Cependant**, l'absence d'un avenir prometteur, étant donné la nature des études suivies connues par le nombre élevé de ses diplômés, constitue elle-même une motivation entrepreneuriale.

(LOCRA-FR – Education)

(121)   White suits, in short, are a matter both of individual taste and personal style. Andrew Marr, one instinctively senses, wisely judges he would carry less credibility in white. Mark Mardell, **on the other hand**, has white-suit cred written all over him.   (Mult-Ed-EN – The Guardian)

(122)   La dette, qui a dépassé la barre symbolique de 1 000 milliards d'euros (63 % du PIB), serait en conséquence sur une pente qui la conduira à 67 % du PIB en 2005, 10 points de plus qu'en 2001. M. Sarkozy a **néanmoins** promis à ses collègues ministres des finances réunis en Irlande de tenir l'objectif de 3 % en 2005. (Mult-Ed-FR – Le Monde)

In Section 8.4.1, it was shown that, whereas CAs used in thematic 1 position tend to perform a strictly linking function, rhematic 1 and rhematic 2 CAs may be used to achieve a range of rhetorical functions in addition to their basic connective function. Based on these observations, it can be postulated that the frequency differences between the two registers may be due to the fact that, while academic prose tends to use conjunctive adjuncts primarily to express purely logical links between discourse units, newspaper editorials more often take advantage of the rhetorical potential afforded by CA placement – and medial positions, in particular – to lay emphasis on a certain part of the message, point the reader towards some of the crucial transitions in discourse, etc. For instance, the analysis of the LOCRA corpus sheds light on a large number of thematic 1 CAs which could very well have been placed in rhematic positions. In (123) and (124), for example, the writers could have made the choice to front the (underlined) topical Theme and use the CA parenthetically right after the adjunct or the subject, in order to lay emphasis on the Theme and the topical shift taking place in their text. Instead, they preferred to place the marker initially, thus establishing it as the point of departure of the message and laying focus on the conjunctive link between the two discourse segments.

(123)   During the first round of disagreements at work, she was diagnosed with depression, which led to further economic penalties and a salary reduction. She filed her first charge of mobbing against the municipality in December 1999, a case that was eventually dismissed. **However**, in her 2003 appeal of her case, the judge found evidence of mobbing, which he defined as a "persecutory strategy in an environment in which the plaintiff was subordinated to a series of behaviors and practices which … damaged, marginalized, and discriminated against her, until it provoked damage to her health" and also to her "moral personhood" (personalità morale). (LOCRA-EN – Anthropology)

(124)   To explore further the distinctive pattern of downward mobility into agriculture, we conducted a discrete-time hazard analysis of the determinants of downward mobility into agriculture for men of rural origins whose fathers worked outside of agriculture. The results showed that the specific kind of nonagricultural occupation of the father had no impact, and neither did father's education or Communist Party membership. **By contrast**, a man's own education is helpful in protecting against mobility into agriculture. (LOCRA-EN – Sociology)

Two main reasons may be advanced for this register difference. Firstly, the greater propensity of newspaper editorials to use CAs rhematically as compared to academic prose may be ascribed to the recent evolution of the register towards a more reader-friendly style of expression. As noted by Cotter (2003: 70) and Westin (2002: 163), the last few decades have witnessed a noticeable rise, in newspaper language in general, and newspaper editorials in particular, in the use of linguistic structures testifying to the reporters' "commitment to their readers […], as a result of economic and social demands made on the news industry to be more 'reader-friendly'" (Cotter 2003: 70). As mentioned at several points in the previous section, CAs placed in rhematic 1 and rhematic 2 positions may be regarded as explicit clues to the reader with respect to the informational and thematic structure of the text. By emphasising the points in discourse at which the topic of the message is shifting – or instead those where it remains constant; by isolating the segments of the message over which the CA has its main scope; or by clearly indicating the boundary between Given and New information, these conjunctive adjuncts help the reader navigate through the main stages of the text. Unlike the editorials, on the other hand, academic prose does not tend to have reader-friendliness among its primary concerns. Rather, as demonstrated by Biber and Gray (2010), some properties of current academic language testify to a fairly high degree of linguistic opacity and implicitness in this register (such as the heavy reliance on phrasal instead of clausal modification, premodifying nouns, appositive noun phrases, etc.), resulting in texts that are not easily accessible to non-expert readers.

In fact, explicitness and reader-friendliness are partly Janus-faced linguistic properties: while they may be viewed as tokens of courtesy towards the reader, facilitating and hence quickening the processing of a given text, they can also be considered as strategies used to orient or constrain the reader's interpretation of that text in a certain way, thus reducing the degree of interpretive freedom that the reader may have towards it. This idea constitutes one of the backbones of Blakemore's (1992: 137) definition of conjunctive markers (or 'discourse connectives', as she calls them): according to Blakemore (ibid.), explicit conjunctive markers may be regarded as devices used by the writer/speaker to "impose constraints on the implicatures of the utterance", i.e. to orient the reader's interpretation of the relationship between two discourse units in a certain way. Likewise, alongside the studies which have underlined the facilitating effect that explicit conjunctive markers may have for discourse processing (see e.g. Degand & Sanders 2002; Sanders & Spooren 2007; see also Chapter 2 on this), some researchers also insist on the 'constraining' character of explicit CMs: for Fraser (1999: 942), for example, conjunctive markers (which he calls 'discourse markers') do not simply display a relationship between two units but rather "impose on S2 a certain range of interpretations, given the interpretation(s) of S1 and the meaning of the DM [discourse

marker]" (see also Hyland 1998; Dafouz-Milne 2008 on the persuasive effect of explicit metadiscourse).[11] It therefore seems as though the boundary between 'guiding' and 'constraining' interpretation is a thin one.

From this it follows that the high frequency of rhematic CAs in the editorial subcorpus may also be attributable to the highly persuasive tone of this register. As already emphasised in Chapter 7, newspaper editorials have been demonstrated to be a very strongly persuasive genre, much more so than academic prose (Biber 1988: 149). Unlike regular news reports, editorials do not chiefly aim to inform the reader about the events that are being discussed since, as explained by Love (2004: 441), editorials typically deal with previously-reported (albeit recent) events, with which the reader is supposed to already be familiar. Instead, the objective is usually to influence the readers' perception and judgement of these events, i.e. to "induce readers to agree with the editor's interpretation and evaluation of recently reported events and situations" (ibid.: 452; see also Bolívar 2002 on this). Interestingly, this implies that the newsworthy information in editorials is not the description of events or situations per se (which are typically conveyed by the Themes of the text, in which they are presented as Given information), but rather the editor's evaluative commentary of those events or situations, which is developed in the rhematic part of the message, and with which the writer attempts to convince the reader to agree (ibid.: 441).

By virtue of their rhetorical potential as emphatic and focusing devices, rhematic CAs are particularly well-suited to the persuasive goals of editorial writing.[12] Placing conjunctive adjuncts in rhematic 1 or rhematic 2 positions may help the reporter lay emphasis on the parts of the message that they view as particularly relevant for their argumentative development. For example, by setting an explicit boundary between Given and New information, rhematic 2 CAs can be used to direct the reader's attention to the writer's evaluative assessment of the event or situation being discussed, hence granting increased attention to the more persuasive portion of the message. This is what happens in Examples (125) and (126). In (125), the information that follows the parenthetical CA *however* expresses the writer's value judgement of the attitude towards securities depicted in the previous

---

11.   Note that these two views do not necessarily stand in contradiction: one may argue that it is by constraining the interpretation and reducing the number of possible readings of a given discourse relation, that explicit CMs facilitate interpretation.

12.   In this respect, it is probably significant that the term 'rhetorical' is defined in the Macmillan Dictionary as "relating to a style of speaking or writing that is effective or intended to influence people" (<https://www.macmillandictionary.com/dictionary/british/rhetorical> (12 April 2020>).

sentence, which s/he considers to be 'misplaced'. By separating it from the rest of the T-Unit, the writer grants this evaluative information increased attention, both with respect to the other elements that surround it, and as compared to the focus that it would have received if the CA had been used in thematic 1 position. In (126), the writer voices his/her opinion on the 'left of the left's' contribution to the French political scene. In the part that follows the contrastive CA *en revanche*, the reporter comments on the limitation of this political trend, viz. that it does not provide the working class with any credible alternatives to the right-wing parties. Although the CA is not separated typographically from the rest of the T-Unit, its occurrence right before the evaluative content of the message contributes to giving this content a somewhat enhanced weight as compared to the previous elements. If one were to read this example aloud, they would probably place a natural (although admittedly light) stress on the word *aucune*, which directly follows the CA. In addition, the rhematic 2 CA of contrast in (126) points the reader's attention towards the part of the message over which it has its main scope (viz. *aucune alternative crédible*, which is contrasted with *un rôle d'aiguillon et de pourvoyeuse d'idées neuves*).

(125)   Securities have a toxic reputation and are firmly associated with the financial crisis. The stigma attached to asset-backed securities is, **however**, misplaced. They have no inherent value.

(Mult-Ed-EN – The Financial Time)

(126)   La gauche de la gauche joue un rôle utile d'aiguillon et de pourvoyeuse d'idées neuves ; elle ne fournit **en revanche** aucune alternative crédible à la droite ou à l'extrême-droite auprès des classes populaires.

(Mult-Ed-FR – Le Nouvel Observateur)

Likewise, rhematic 1 CAs also provide the writer with the opportunity to grant focus to the part of the message which they deem particularly important for their argumentative development. In (127), for instance, the rhematic 1 use of the CA contributes to reinforcing the writer's perception of the position of EU members towards the Ukrainian crisis: by setting the Theme (viz. 'the most immediate dilemma') apart from the rest of the sentence, the writer gives it even more weight than it already carried by virtue of its mere ideational content (and more particularly through the use of a superlative). Likewise, in the fairly long extract reproduced in (128), the emphasis of the subject pronoun *aucune* by means of rhematic 1 *cependant* strengthens the power of generalisation of the argumentation that follows: while in S1, the writer acknowledges that *all* of the criticisms are founded, in S2 s/he insists that they are also *all* questionable, in three main respects. In (129), too, the rhematic 1 use of *however* reinforces the writers' position on the project of a shake-up of Britain's welfare system. More precisely, it stresses the writer's view that, while it poses a number of problems, this reform is

absolutely *not* a luxury. In this case, the effect is further strengthened by the choice of topical Theme, which takes the form of a so-called 'thematic equative' (Halliday & Matthiessen 2004: 69). Thus, in all three examples, the rhematic CA contributes to reinforcing the writer's argumentative development, by emphasising some of the most important aspects of their position towards the issues being discussed.

(127)   Since the crisis over Ukraine first erupted, EU member states have wavered over how much economic pain they are prepared to inflict on themselves to contain Russian aggression. The concerns are various. Britain fears that the imposition of financial and banking sanctions would lead to a rush of Russian capital out of the City of London. Germany and Italy worry that an energy embargo would prompt Russia to stop the supply of oil and gas on which they rely. <u>The most immediate dilemma</u>, **however**, is faced by France, which is in the final stages of selling two warships to Russia for EUR1.2bn.
(Mult-Ed-EN – The Financial Time)

(128)   La diversion, d'abord. La contre-attaque du chef de l'Etat, a souligné l'opposition, ne serait qu'un écran de fumée pour tenter de faire oublier sa part de responsabilité dans l'"affaire Cahuzac" : négligence ou naïveté lors de la nomination de cet homme comme ministre du budget, crédulité ou aveuglement depuis qu'il avait été mis en cause, en novembre 2012. L'inquisition, ensuite. En proposant, notamment, de rendre obligatoire la publication du patrimoine des "principaux responsables politiques et administratifs du pays", au premier rang desquels les ministres et les parlementaires, M. Hollande céderait, selon ses détracteurs, à la dictature de la transparence, avec le risque que cela peut comporter pour les libertés individuelles. Pis, jetant ainsi l'opprobre sur tous les élus, il les livrerait en pâture à l'opinion publique, à son voyeurisme et au populisme ambiant. L'improvisation, enfin. Préparées dans l'émotion et l'urgence, les mesures annoncées relèveraient du bricolage et seraient, au bout du compte, inefficaces contre des fraudeurs déterminés. En outre, l'interdiction annoncée de cumuler un mandat parlementaire avec une activité professionnelle, sauf exceptions, soulève d'infinis problèmes. Aucune de ces critiques n'est sans fondement. <u>Aucune</u>, **cependant**, ne résiste à trois évidences. D'une part, l'on peut faire confiance au Conseil d'Etat en amont, aux parlementaires eux-mêmes lors de l'examen des lois en préparation, puis au Conseil constitutionnel en aval, pour éviter des dispositions par trop contestables ou impraticables. D'autre part, l'on ne peut ignorer que la plupart des pays démocratiques, en Europe et au-delà, ont instauré des règles de transparence et de contrôle beaucoup plus sévères qu'en France […] Enfin, et surtout, il faudrait savoir ce que l'on veut. Ou bien l'on estime que la crise morale révélée par l'affaire Cahuzac est très grave et que la défiance des citoyens à l'égard de leurs élus (considérés comme "corrompus" par les deux tiers des Français, voire davantage) est insupportable, et l'on se

donne alors les moyens, rigoureux et vigoureux, de lutter contre ce cancer
pour la démocratie. Ou bien, pour mille prétextes, l'on renonce à cet effort
de moralisation   (Mult-Ed-FR – Le Monde)

(129)   A shake-up of Britain's costly, Byzantine and often unfair welfare state
would be painful in any event; at a time of economic upheaval, recession
and a straitened Treasury, it is trickier still. <u>What it is not</u>, **however**, is a
luxury.   (Mult-Ed-EN – The Independent)

Finally, as demonstrated in Section 8.3.2, the rhematic 1 and the rhematic 2 positions are each preferred by one language: rhematic 1 CAs are more common in English than in French, which displays a very strong preference for rhematic 2 markers. As these two positions were demonstrated to perform partly different types of discourse effects, these frequency discrepancies may reflect differences in the ways in which English and French handle thematic and information structure, and in the rhetorical strategies that English and French writers use in order to fulfil persuasive communicative functions. However, it should be noted that from the type of analyses performed here, it was not possible to clearly disentangle the parts played by the typological differences between languages (e.g. with respect to the reluctance of English to interrupt the SVO structure) and the preferred rhetorical strategies of the two languages. Further research is therefore required in order to tease out the respective influence of these two factors on the differences in frequency of rhematic 1 and rhematic 2 CAs between the two languages.

8.5   Conjunctive adjunct placement at the syntax-lexis interface

So far, this chapter has provided a fairly broad account of conjunctive adjunct placement: the category of CAs of contrast has been viewed as a relatively homogeneous one, and no account has yet been taken of the possibility that the various lexical items within each language system may display different placement preferences. Yet, as explained in Section 8.1, there is some evidence in the literature that conjunctive adjuncts do not always display the same syntactic behaviour: with respect to English, researchers such as Jacobson (1964), Altenberg (2006) or Paquot (2010) have shown that some conjunctive adjuncts display partly idiosyncratic placement patterns. In this section, I attempt to evaluate – and compare – the role played by lexis on CA placement in each language system. The analyses presented in this section thus move in the opposite direction to those presented in Section 8.4: whereas the previous section was devoted to the exploration of the relationship of syntax with a higher level of the linguistic system (viz. discourse), the present section focuses on the ties between CA placement and a lower level of the system, viz. lexis. The first part of the analysis compares the placement patterns of the various conjunctive adjuncts in each language, irrespective of register

variation (Section 8.5.1). The objective is to answer the third major research question asked in this chapter, i.e. to what extent does lexis influence the positional patterns of English and French conjunctive adjuncts. In a second stage, the individual placement profiles of the CAs are compared in each register, in an attempt to test whether the extent of the differences between the various lexical items is influenced by register variation. In other words, Section 8.5.2 focuses on the fourth and last research question asked in this chapter, viz. do lexis and register interact to influence CA placement, and if so, in what ways?

8.5.1   Individual placement patterns of English and French conjunctive adjuncts of contrast

The percentages of occurrences of each conjunctive adjunct in each of the five positions are provided in Figure 28 for English, and Figure 29 for French. The CAs are arranged in decreasing order of frequency of occurrence in thematic 1 position, and the raw frequencies of occurrence of each marker in the corpus are provided between brackets below each bar of the plot. For the sake of visual clarity, the percentage values which do not reach 5% are not displayed on the graphs.

8.5.1.1   *English*
As can be seen from Figure 28, the results obtained for English strongly confirm the findings reported in previous studies on CA placement (see Jacobson 1964; Altenberg 2006; Paquot 2010): they provide evidence of clear lexical variation in placement between the eight conjunctive adjuncts. The most strikingly idiosyncratic CA in this respect is *though*, which is used in thematic 1 position markedly less frequently than the other markers (3.7% of the occurrences). The fact that *though* should display even a few instances of the thematic 1 position is actually surprising in itself: as already mentioned earlier, the clause-initial position is usually reserved to subordinating uses of *though*, as in (130). In this respect, it is important to note that none of the thematic 1 instances of adverbial *though* were purely initial: rather, they all followed another textual Theme, as in Example (131).

(130)   **Though** the new list includes a handful of people of real distinction who will adorn the second chamber, the wheat is greatly outnumbered by the chaff: superannuated politicians, party or trade union officials, donors and courtiers.   (Mult-Ed-EN – The Telegraph)

(131)   How did Chancellor Merkel react to the demands put forward by her defense minister and foreign minister respectively? At first glance, it appears that she had little choice but to accept the criteria put forward by the two. Otherwise, she would have damaged the credibility of her defense minister, stirred conflict among the coalition parties by challenging her foreign minister, or called into question the ability of her government to

make important foreign policy decisions in general. <u>In addition</u>, **though**, the conditions were also useful to Merkel with respect to achieving her own policy preferences.   (LOCRA-EN – Political science)

In addition, *though* also displays a markedly higher frequency of use in rhematic 1 position (72.3%) – as in Example (132) – than the other seven English CAs. This result is also very surprising, since this marker has frequently been claimed to be strongly associated with the sentence-final position (e.g. Quirk et al. 1985: 643; Altenberg 1986: 22; Lenker 2010: 208). Yet, while *though* is indeed the CA with the highest proportion of rhematic 3 occurrences in the corpus (10.6%) – as exemplified in (133) – this position remains far (viz. seven times) less common than the rhematic 1 slot. One possible explanation for this may be that the statements formulated on the position of adverbial *though* have so far been founded mainly on the analysis of spoken data. Thus, in this specific case, mode would also appear to play a significant role on CA placement.

(132)   The Bigley family, whose hopes and fears have been cruelly manipulated by the terrorists, deserves all our sympathy. Their anger, **though**, should be directed not at Tony Blair, but at the murderers.   (Mult-Ed-EN – The Telegraph)

(133)   Bound as he is by collective cabinet responsibility, the Business Secretary is wary enough of the term "bubble", let alone of any suggestion that the Coalition is doing its share of pumping. It most certainly is, **though**.

(Mult-Ed-EN – The Independent)
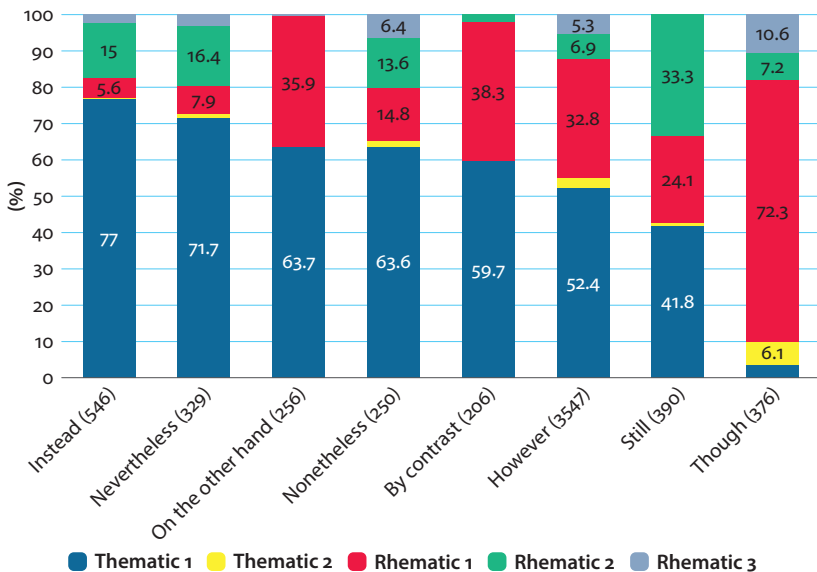
Figure 28. Individual placement patterns of English conjunctive adjuncts of contrast (in percent)

Alongside the strikingly idiosyncratic positional behaviour of *though*, the results presented in Figure 28 also provide further evidence of the influence of lexis on CA placement. For example, even when excluding *though* from the comparison, and although this position remains the most frequent one for all the markers, there is a 35% range of variation in proportion of use of the thematic 1 position across the seven CAs, with values stretching from 41.8% (for *still*) to 77% (for *instead*, see Example (134)). Likewise, still excluding *though*, rhematic 1 CAs are much more strongly associated with some CAs than others: for example, while *by contrast* occurs in rhematic 1 position in 38.3% of the cases, as in (135), this position only accounts for 5.6% of the occurrences of *instead*. In addition, the corpus results display a substantial amount of variation between these two extremes (e.g. 14.8% for *nonetheless*; 24.1% for *still*). The proportion of rhematic 2 uses, too, ranges from 0% for *on the other hand*, to a third of the CA tokens for *still*, as in (136). Finally, Figure 28 also highlights some variation in the amount of positional mobility of the markers: for example, whereas *on the other hand* and *by contrast* are (almost) restricted to the thematic 1 and rhematic 1 positions, and both seem to avoid the rhematic 2 position, *still* is nearly evenly distributed between these three positions. Similarly to *though*, the positional patterning of *still* is actually very surprising when compared to the descriptions that are usually made of this CA in the literature. *Still* is generally cited by descriptive grammars as one of the markers which are virtually restricted to sentence-initial position (e.g. Quirk et al. 1972: 444; 1985: 643). Yet, the corpus results show that *still* occurs in thematic 1 position in about 40% of the cases: in fact, among the eight English CAs investigated here, *still* is the second least frequent marker in this position (after *though*). In addition to its sentence-initial uses, *still* also exhibits a high frequency of occurrences in rhematic 2 (33.3%) and rhematic 1 (24.1%) positions, as in Examples (136) and (137), respectively.

(134)   My aim here is not to offer a total account of neoliberalism(s), if such a thing were possible (Hoffman et al. 2006; Kingfisher and Maskovsky 2008; Ong 2006; although see Comaroff and Comaroff 2000; Gershon 2011; Gowan 1999; Harvey 2005 for useful synoptic discussions). **Instead**, I want to focus on a number of socioeconomic changes, discussed under the rubric of neoliberalism, in order to outline some of the ways in which the brand and its surfeits have functioned as mediating technologies of and for neoliberalism.   (LOCRA-EN – Anthropology)

(135)   Twenty-one years ago, after Michael Fagan got into the Queen's bedroom, William Whitelaw immediately offered to resign. Mrs Thatcher rightly turned him down. Mr Blunkett, **by contrast**, has responded to the latest penetration of Buckingham Palace security by denying any responsibility whatever - even insinuating to MPs that the Queen supports his denial.
                                                           (Mult-Ed-EN – The Guardian)

(136)   The justification offered for the war by Mr Blair may have been the wrong one, but it was **still** a just war.   (Mult-Ed-EN – The Telegraph)

(137)   Mr Blair could provide valuable help to MPs as they work their way though [sic.] some crucial questions. When did this scheme begin, how did it operate, was proper advice listened to, what were the numbers involved and when did it end? Leaders in the peace process may have found themselves making decisions they were far from comfortable with, and the public will probably understand that. But the victims **still** deserve honesty.

(Mult-Ed-EN – The Telegraph)

8.5.1.2   *French*

As appears clearly from Figure 29, French conjunctive adjuncts also display partly idiosyncratic placement patterns. As was the case in English, one marker in particular, viz. *plutôt*, stands out by exhibiting a very low frequency of occurrence in thematic 1 position as compared to the other six markers. Rather, nearly nine occurrences of *plutôt* out of ten in the corpus are used within the predicate, as in (138) and (139).

(138)   Dans les travaux décrits précédemment, on modélise les opportunités offertes aux apprenants de simplifier les informations portées par les stimuli afin de les catégoriser en utilisant une somme d'information minimisée. Les travaux suivants focalisent **plutôt** sur la possibilité de changer de règle en utilisant de nouvelles informations.   (LOCRA-FR – Psychology)

(139)   Surtout, rien ne prouve que Sharon réussira à terroriser les terroristes. De l'aveu même des services de renseignements israéliens, le Hamas n'est pas une pyramide dont il suffirait de couper la tête pour la liquider. Son organisation en cellules rappelle **plutôt** la grappe de raisin, chaque grain vivant en autonomie.   (Mult-Ed-FR – Le Figaro)
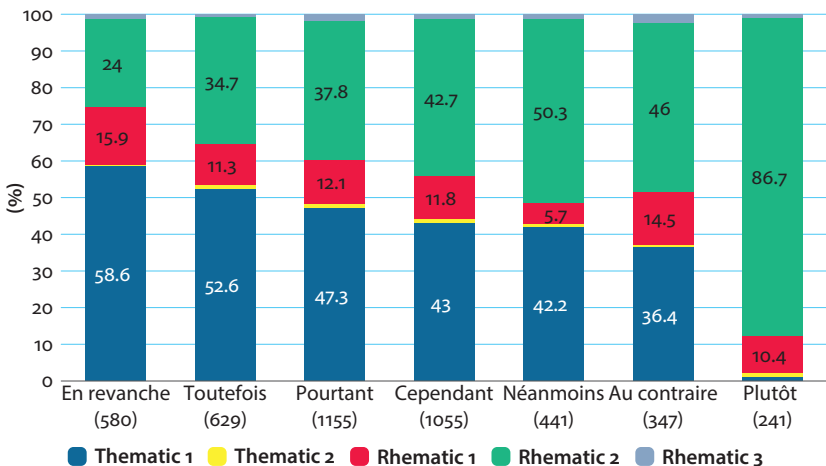
Figure 29. Individual placement patterns of French conjunctive adjuncts of contrast (in percent)

Other striking lexical differences emerging from Figure 29 pertain to the proportions of use of the thematic 1 and rhematic 2 positions across CAs. Firstly, as opposed to what was observed in English, the various CAs partly differ with respect to their preferred position: whereas *en revanche, toutefois* and *pourtant* are most commonly used sentence-initially, the preferred position of *au contraire* and *néanmoins* is the rhematic 2 slot. The occurrences of *cependant*, on the other hand, are roughly equally distributed between the two categories. The proportions of thematic 1 and rhematic 2 occurrences thus vary quite noticeably across markers – although these two positions are arguably important for all the CAs. Excluding the very peculiar *plutôt*, there is a range of about 25% variation in use of each of these positions across the six CAs, viz. between 36.4% (*au contraire*) and 58.6% (*en revanche*) for the thematic 1 position – exemplified in (140); and between 24% (*en revanche*) and 50.3% (*néanmoins*) for the rhematic 2 position – as in Example (141).

(140)   Le lien avec les contenus de cours faisait à chaque fois l'objet d'explications détaillées. **En revanche**, peu de références au contenu des tweets étaient faites en cours.   (LOCRA-FR – Education)

(141)   Dans la mesure où les traitements contre l'esca sont ponctuels et s'étalent sur deux à trois jours d'hiver par an, les risques de contamination des travailleurs agricoles utilisant cette substance furent jugés négligeables. Les expositions à l'arsenite de soude furent **néanmoins** soumises à un contrôle spécifique de la part de la Mutualité sociale agricole.
                                            (LOCRA-FR – Anthropology)

Note, however, that this range of variation is slightly more limited than those observed for the most frequent positions in English, viz. thematic 1 and rhematic 1 (c. 35% in both cases). Likewise, the amount of variation observed with respect to the rhematic 1 position is much more restricted than was the case in English, since most French CAs display a similar proportion of rhematic 1 uses (around 10%). More generally – and still excluding *plutôt* – if the colour pattern of Figure 29 is compared with that of Figure 28, it appears that the positional patterning of French CAs is rather more homogenous than that of their English counterparts, with a general dominance of the rhematic 2 (green) and thematic 1 (dark blue) positions, a few instances of rhematic 1 uses, and very few examples of either the thematic 2 or the rhematic 3 positions. In Figure 28, on the other hand, the colour repartition in each bar of the plot is more uneven across CAs. Finally, unlike Figure 28, Figure 29 does not appear to reveal any striking differences in positional flexibility between the markers: all CAs appear to be perfectly acceptable in thematic 1, rhematic 2 and, to a lesser extent, rhematic 1 positions (except, again, for *plutôt*, which avoids the thematic 1 position).

In summary, the study of both English and French data provides conclusive evidence of lexical variation in placement between the conjunctive adjuncts

within each language system. The results obtained here thus provide another example of lexical priming (Hoey 2005; see Chapter 7, Section 7.3.3 for a definition) with respect to conjunctive markers of contrast: in the same way as the syntactic patterning of conjunctive markers of contrast was found to vary lexically (see Chapter 7), so do the placement preferences of English and French conjunctive adjuncts. As already explained earlier in this chapter, descriptions of conjunctive adjuncts in the literature tend to formulate fairly general rules of placement for languages as wholes, usually accompanied by the recognition that these items are nevertheless syntactically mobile, and may therefore appear in various positions in the sentence. When lexis-sensitive comments are available, they tend to be restricted to a few well-known cases, such as the strong preference of *still* for initial position, or that of *though* for final position (e.g. Quirk et al. 1972: 444; 1985: 643; Biber et al. 1999: 891–892). With respect to the influence of lexis, the corpus findings do not strictly call into question the relevance of formulating general rules of CA placement for a given language: despite some amount of lexical variation, the thematic 1 slot was broadly dominant in English, while most French markers were commonly found either in thematic 1 or rhematic 2 position.[13] The corpus findings presented in this section may nevertheless be considered as a valuable complement to the general statements found in the literature, for two reasons. Firstly, the results have shown that the lexis-sensitive comments made in the reference literature are not necessarily supported by empirical evidence (or at least, evidence obtained on the basis of written data), since the results obtained for both *still* and *though* largely contradicted the placement guidelines found in the literature. Secondly, it may be useful to raise users' awareness to the fact that not all lexical items are equal when it comes to either (i) their propensity to resort to the position(s) presented as dominant in the literature (since substantial variation in use of these positions was found across lexical items, in both English and French); or (ii) their openness to the other positional alternatives (viz. medial or final) presented in these books. For example, it is important that users be made aware that, although *however, by contrast* or *on the other hand* work very well in rhematic 1 position (with each CA occurring in this position around 35% of the time), *instead* and *nevertheless* should only be used parsimoniously between the topical Theme and the verb phrase (where they occur less than 10% of the time).

---

13.   It is nevertheless worth remembering that the limitations of such general placement rules were clearly underlined in the cross-register analysis of English and French CA placement carried out in Section 8.3.

8.5.2   The combined influence of lexis and register on conjunctive adjunct placement

The previous section provided evidence that in both French and, especially English, different conjunctive adjuncts display partly distinct placement preferences. In this section, the influence of lexis is investigated in conjunction with that of register. The objective is to assess whether the type of influence exerted by lexical variation on CA placement differs across communicative situations. The individual placement patterns of each conjunctive adjunct per register are provided in Figure 30 for English, and Figure 31 for French. The raw frequencies of occurrence of each marker in each register are provided between brackets. Again, for greater readability, the percentage values that do not reach 5% are not displayed on the graphs.

8.5.2.1   *English*

As Figure 30 shows, investigating CA placement from the angle of both lexis *and* register provides additional insights to those that may be gained when analysing each factor separately. Firstly, the graph indicates that there appears to be slightly more lexical variation in the editorial than in the academic register: whereas in academic prose, the thematic 1 position largely dominates across markers, with all but two markers (viz. *still* and *though*) displaying over 60% of their occurrences in this position, the picture is slightly more varied in the editorials, where the rhematic 1 position dominates for some CAs (viz. *however* and *by contrast*). In fact, what Figure 30 shows is that some conjunctive adjuncts display distinctly different placement preferences depending on the register. If we look at *by contrast*, for instance, we observe a nearly perfectly inverted ratio of occurrence of the marker in thematic 1 and rhematic 1 positions: whereas in the research articles, *by contrast* is mostly found in thematic 1 position (78.4% of the tokens) – as in (142), its preferred position in the editorials is rhematic 1 (60%) – as in (143). In addition, whereas in the academic subcorpus, *by contrast* is restricted to the thematic 1 and the rhematic 1 positions, in the editorials, the CA also appears (albeit more rarely) in rhematic 2 position, as in (144).

(142)   This presentation of the data makes clear that four topics are more common in the media than in the oral questions: justice, rights, defence and government (such as elections). **By contrast**, members of parliament spend considerable time focusing on the routine matters of public policy such as transport, health, energy and foreign trade, all of which are less interesting to the media, or at least are relatively rarely seen in the front pages.

(LOCRA-EN – Political science)
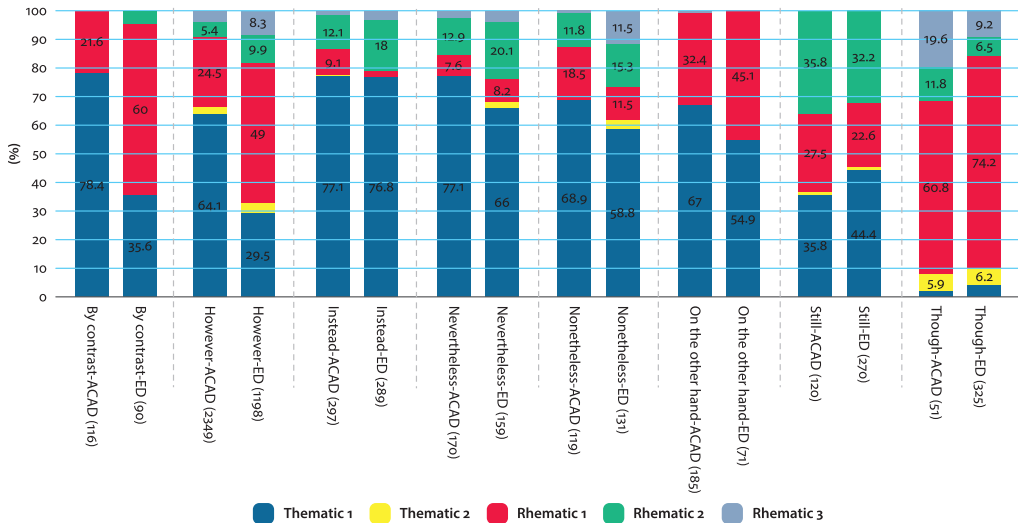
Figure 30.   Individual placement patterns of English conjunctive adjuncts of contrast per register (in percent)

(143)   Last month, to Beijing's annoyance, several pro-democracy figures in Hong Kong testified before the Senate foreign relations committee in Washington. The British Government, **by contrast**, has been reluctant to offend the Chinese.   (Mult-Ed-EN – The Telegraph)

(144)   It may be that those testimonies already exist among the documents that the Home Office and police did manage to retain. The inquiries will surely make these their priority. The inquiries announced by Theresa May will not, **by contrast**, clear the names of those named in lurid web conspiracy theories.   (Mult-Ed-EN – The Independent)

Likewise, whereas the preferred position of *however* in LOCRA is clearly thematic 1 (64.1% of the occurrences) – as in (145); in Mult-Ed this position only accounts for about 30% of the occurrences. Rather, the most common patterning of *however* in the editorials is the rhematic 1 use – as in (146) – which accounts for half of the occurrences of the marker in this register.

(145)   Simply put, without the availability of racially diverse peers, interactions across race are literally impossible (Blau & Schwartz, 1984). **However**, students' experiences with CRI are also influenced by a myriad of other factors.   (LOCRA-EN – Education)

(146)   Taxes on investment income in Italy are generous by European standards. That the extra revenue levied on savings will be used to give some breathing space to businesses should help to boost growth. This cash, **however**, will only fund a portion of the promises that Mr Renzi has made.
                                              (Mult-Ed-EN – The Financial Time)

Other markers, by contrast, appear to display relatively stable positional preferences across registers. This is the case of *instead,* for example, which displays 77% of thematic 1 uses, whatever the register in which it occurs. *Still* also displays remarkably similar placement patterns across registers: in both LOCRA and Mult-Ed, the occurrences of *still* are distributed across the thematic 1, rhematic 2 and, less frequently, rhematic 1 positions. *Though* and *nevertheless*, too, have relatively similar placement patterns in the two registers, although some categories are slightly more common in one register than the other (e.g. rhematic 2 *nevertheless* and rhematic 1 *though* are more common in the editorials).

8.5.2.2   *French*

Figure 31, which displays the individual patterning of French conjunctive adjuncts of contrast per register, also provides additional insights to those gained from the separate analyses of lexis and register. As was the case in English, the plot reveals that the positional preferences of some conjunctive adjuncts are sensitive to register variation. A case in point is *cependant*: while more than half (52.8%) of the
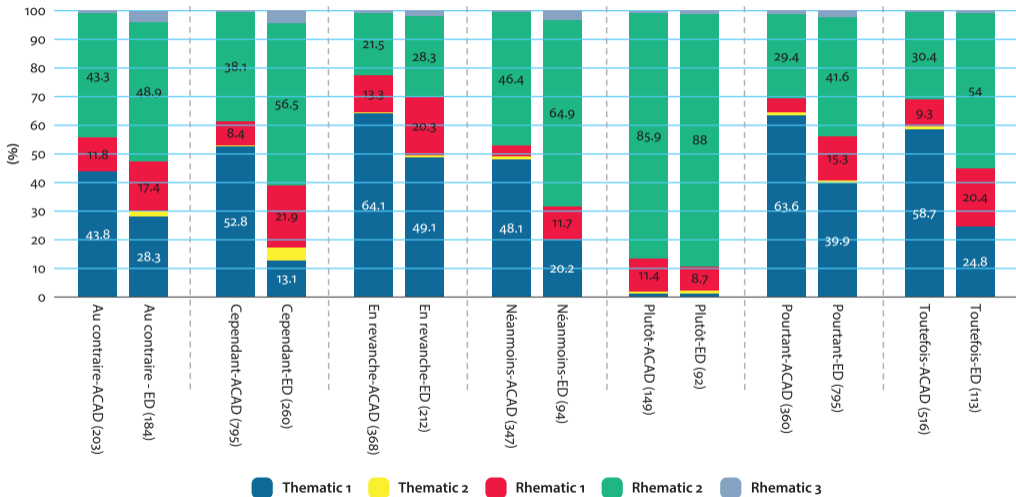
Figure 31. Individual placement patterns of French conjunctive adjuncts of contrast per register (in percent)

occurrences of this marker are used thematically in the academic register – as in (147), in Mult-Ed this proportion drops to only 13.1% of the occurrences. Instead, the rhematic 2 category represents the largest proportion of the data (56.5%) in the editorials – as in (148).

(147) Les manifestations d'entités tutélaires peuvent suivre des modalités culturellement définies ou être simplement constatées, certains animaux, objets et personnes étant privilégiés par les différentes déités ancestrales. **Cependant**, aucune relation fixe et permanente n'est ici reconnue entre une entité et le réceptacle qui, éventuellement, l'accueille, comme l'attestent les tombes qui offrent aux défunts le moyen de s'ancrer dans le monde des vivants et d'interagir avec eux.   (LOCRA-FR – Anthropology)

(148) Les commandants de bord qui le désirent ont désormais également le droit de porter une arme à feu dans la cabine de pilotage. Il leur suffit d'effectuer un stage de deux semaines. Beaucoup l'ont fait, apparemment. Ils n'ont **cependant** pas le droit d'utiliser leur arme hors du cockpit.
                                                        (Mult-Ed-FR – Le Monde)

The same pattern is observed for *néanmoins, toutefois* and *au contraire*, for which the proportion of thematic 1 uses – as in Examples (149) to (151) – is systematically markedly higher in the academic than in the editorial register, where other positions are also frequent. For example, rhematic 2 uses of *néanmoins* and *toutefois* – as in (152) and (153) – are far more frequent in the Mult-Ed than in the LOCRA subcorpus, as are the rhematic 1 instances of these CAs, exemplified in (154) and (155).

(149) En proposant une prise en charge de la malnutrition au niveau de la famille, ces nouveaux usages des ATPE constituent une innovation radicale qui permet d'éviter l'hôpital et ses contraintes multiples. **Néanmoins**, ce nouveau modèle doit faire face à des réticences fortes qui vont conduire à considérer les premières expériences de déploiement de ces produits dans le cadre d'une prise en charge ambulatoire comme pouvant mettre en danger la santé des patients.   (LOCRA-FR – Anthropology)

(150) On remarque tout d'abord que le taux de changement n'est pas lié à la diversité du conseil présent, puisque globalement ces changements sont du même ordre dans les deux communes (Tableaux 3 et 4). **Toutefois**, dans chacune des communes, certains viticulteurs bénéficient plus spécifiquement d'un type de conseil que d'un autre.
                                                        (LOCRA-FR – Sociology)

(151) Le père n'a jamais demandé aux deux plus jeunes sœurs de lui venir en aide. **Au contraire**, il semble conserver Jimmy dans son rôle d'ainé malgré sa situation de handicap.   (LOCRA-FR – Education)

(152)   La mission recommande sagement aux pouvoirs publics de se concentrer sur les contrevenants à grande échelle, notamment ceux qui cherchent à faire argent de l'illégalité. Elle prévoit **néanmoins**, comme aujourd'hui, une "riposte graduée" et des amendes en cas de persistance du comportement fautif.   (Mult-Ed-FR – Le Nouvel Observateur)

(153)   Pour priver Bayrou d'oxygène, Sarkozy doit en offrir aux UDF qui l'ont rallié. Et faire vivre une vraie droite plurielle. Depuis des années, il n'imagine **toutefois** son camp qu'unanime et dévoué à sa personne.
                                              (Mult-Ed-FR – Libération)

(154)   Mais ce serait oublier que la Toile n'est ici qu'un messager et que, dans ce cas précis, les deux adolescentes ont également correspondu par petites notes et coups de téléphone. Le Web, **néanmoins**, porte la responsabilité qui est la sienne : celle d'un média hyperpopulaire parmi les plus jeunes pour communiquer entre eux et qui peut toucher une audience beaucoup plus large que n'importe quelle lettre désespérée.   (Mult-Ed-FR – Libération)

(155)   Vladimir Poutine n'a pas de complexes avec l'histoire. Il vient de comparer le siège de Donetsk à celui de Leningrad par les troupes nazies durant la Seconde Guerre mondiale. On voit quel esprit de résistance il veut ici appeler. La comparaison, **toutefois**, semble un brin excessive.
                                              (Mult-Ed-FR – le Figaro)

On the other hand, also in line with what was observed in English, some French markers display relatively stable placement patterns across registers. The most striking example of this is *plutôt*, whose placement preferences are nearly identical in LOCRA and Mult-Ed. The fact that the positional patterning of *some* CAs appears to be influenced by register variation, while that of others tends to remain stable across registers may suggest that lexis and register do not influence CA placement fully independently of each other, but rather that there is a certain degree of interaction between these two variables with respect to placement: whereas for some lexical items, register plays a part in the placement choices made by the writers, other CAs seem to be more impervious to register variation.

8.5.2.3   *Respective influence of lexis and register on English and French conjunctive adjunct placement*

In order to gain deeper insights into the respective impact of lexis and register on conjunctive adjunct placement, and the nature of the relationship between these two factors of influence, the statistical method of Classification and Regression Trees was once again applied to the data. The objective was to predict CA placement, no longer from language and register, but from the register and the lexis variables. Unfortunately, it was not possible to build a predictive model integrating the three factors simultaneously (viz. language, lexis and register), since such a model would have required establishing one-to-one equivalences between the

various levels of the 'lexis' variable in each language (e.g. by saying that English *however* corresponds to French *cependant*, that *by contrast* is equivalent to *en revanche*, etc.). As already made clear at several points in this book, it is usually not possible to establish clear, one-to-one mappings between CMs across languages (see Chapter 3 for a detailed discussion). Therefore, the respective influence of lexis and register is assessed separately for each language: the CART model for English is provided in Figure 32, while the model for French CAs is given in Figure 33.

8.5.2.3.1   English

As Figure 32 shows, with respect to English CA placement, the first distinction operated by the CART method pertains to lexis. The model shows that, whatever the register, the placement preferences of *though* differ markedly from those of the other seven markers. More specifically, *though* differentiates itself from the other CAs by its strong preference for the rhematic 1 position. In the rest of the data set, on the other hand, the thematic 1 position is the one that prevails. Once *though* has been set apart from the rest of the data set, the next best predictor of CA placement identified by the model is register. The CART method reveals that, while in the academic register all CAs behave in roughly similar ways, with a propensity to occur in thematic 1 position, the preferred placement patterns of CAs in the editorial register vary lexically: whereas *instead, nevertheless, nonetheless* (along with its orthographic variant
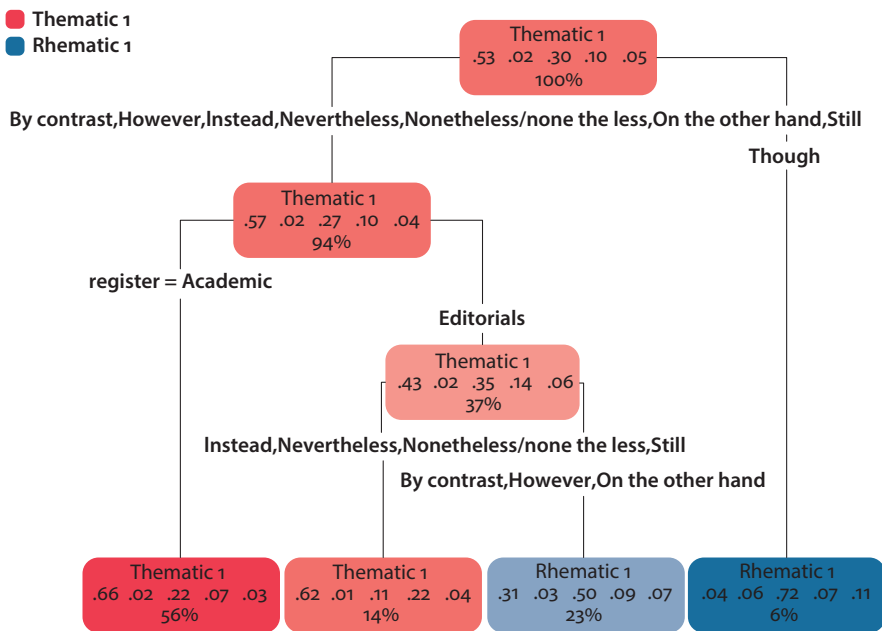

Figure 32. Classification and Regression Tree for conjunctive adjunct placement as a function of register and lexis in English

none the less) and *still* tend to occur in thematic 1 position, *by contrast, on the other hand* and *however* display a special preference for the rhematic 1 position. In other words, the model confirms the existence of an interaction between lexis and register with respect to CA placement, since the amount of influence exerted by lexis on CA placement depends on the register in which these markers are used (see also Berk 2016: 133 on the identification of interaction effects in Classification and Regression Trees). Whereas some lexical differences are uncovered in the editorials, no such distinctions emerge from the academic subcorpus – except for *though*, which differs from the other markers irrespective of register. This suggests that, in order to provide a comprehensive account of CA placement in a given language, it is necessary to look at the influence of both lexis and register simultaneously.

Looking at the data from another perspective, Figure 32 also confirms the observation that, while some CAs display stable placement patterns across registers, the placement preferences of other markers appear to be sensitive to register variation. The graph clearly shows that, whereas *though, instead, nevertheless, nonetheless* and *still* tend to display a fixed placement profile across registers (i.e. they show a preference for the thematic 1 category – or rhematic 1, in the case of *though* – in both registers), the syntactic preferences of *by contrast, however* and *on the other hand* are not the same in the research articles (thematic 1) and the editorials (rhematic 1). In other words, two main types of placement profiles emerge from the corpus analysis: on the one hand, some CAs may be said to display strong lexical priming (cf. Hoey 2005; see Chapter 7 for a definition), in the sense that lexis seems to be the factor that has the more influence on their position. Whatever the communicative situation in which they are used, they display roughly the same placement tendencies. It is clear, however, that the present analysis only investigates written data, and that the pattern of stability uncovered here is probably not generalisable to all communicative situations. As already mentioned earlier, for example, *though* would be unlikely to show such a strong preference for the rhematic 1 position in a corpus of spoken data. Other conjunctive adjuncts, on the other hand, appear to display what we might call 'stylistic priming': their placement preferences vary depending on the communicative situation in which they are used.

8.5.2.3.2   French

The CART tree emerging from the analysis of the French data (Figure 33) displays a remarkably similar picture to that obtained for English. Again, the first division made by the model pertains to lexis: the placement patterns of *plutôt* are identified as distinctly different from those of the other six markers. Unlike the other CAs, *plutôt* tends to prefer the rhematic 2 position, whether it is used in the editorial or in the academic register. Still in line with the results obtained for English, the CART model then indicates that the next best predictor of CA

placement in French is register. Again, register is the factor that is going to predict the amount of lexical variation found in the data. In academic prose, all markers (excluding *plutôt*) are shown to behave in roughly similar ways (i.e. they all display a clear propensity to occur in thematic 1 position), as is clear from the fact that the model makes no further distinctions between lexical items within this register. By contrast, in the editorials, some CAs (viz. *en revanche, pourtant*) appear to display a preference for the thematic 1 position, whereas other markers (viz. *au contraire, cependant, néanmoins* and *toutefois*) are more common in rhematic 2 position. Thus, as was the case in English, the data reveals a certain degree of interaction between lexical and stylistic factors with respect to placement: the role played by lexis partly depends on the register in which the markers are used.
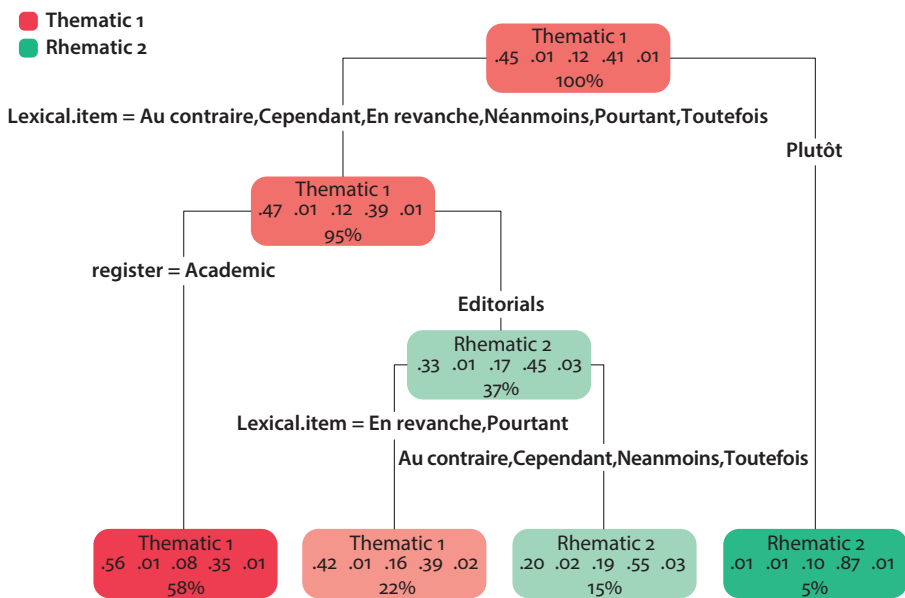
Figure 33. Classification and Regression Tree for conjunctive adjunct placement as a function of register and lexis in French

The French data also matches the distinction between lexically-primed and stylistically-primed conjunctive adjuncts made above. The corpus results indicate that, alongside markers which display fixed placement patterns across registers (e.g. *en revanche, pourtant* are associated with the thematic 1 slot in both registers; *plutôt* occurs in rhematic 2 position independently of the type of text being analysed), some CAs exhibit variable placement preferences across registers. Therefore, depending on the situational context, *au contraire, cependant, néanmoins* and *toutefois* will show a preference either for the thematic 1 position (in academic prose) or the rhematic 2 slot (in the editorials).

In summary, the investigation of CA placement through the lens of both lexis and register has shed light on the sheer complexity of the phenomenon. The results demonstrated not only that both lexical and register factors play a role on CA placement choices, but that these two variables interact in influencing placement. In comparison, no such interaction was found when the respective weight of linguistic and register factors was assessed (see Section 8.3.3): rather, that part of the analysis showed that CA placement was influenced first by language, then by register, with each factor affecting CA placement rather independently of the other one. As regards lexis and register, it is probably more complex to determine with any certainty which factor is the more influential one. One way of looking at the phenomenon may be to say that the respective weight of each variable seems to depend on the lexical item: whereas the placement patterns of some CAs are primarily influenced by register – hence being 'stylistically-primed' – the position of other, 'lexically-primed' markers is determined solely by lexis. Alternatively, we could also say that the amount of lexical variation depends on the register: whereas marked variation in placement is observed in the editorial subcorpus, CA placement is more stable in the research articles.

Interestingly, in the light of the evidence presented earlier in this chapter, these two types of placement profiles may be associated with partly diverging degrees of polyfunctionality of the markers. As was demonstrated in Section 8.4, CAs occurring in rhematic positions tend to perform a range of discourse functions in addition to their primary linking function. In view of this, we may postulate that CAs displaying variable placement patterns across registers (viz. occurring mostly in thematic 1 position in LOCRA, but in rhematic position in Mult-Ed) might have a greater capacity than the other markers to create specific discourse effects in accordance with the communicative goals of the register in which they are used (viz. here, newspaper editorials). In other words, while some markers remain bound to the thematic 1 position whatever the context, others may move to rhematic position in registers whose communicative purposes (e.g. persuasiveness or reader-friendliness) require it.

Finally, it is very striking that lexis and register appear to influence placement in strikingly similar ways in English and French: in both languages, the CART method returned the same tree, with (i) one item (viz. *though, plutôt*) being set apart from the others right from the start due to a highly idiosyncratic placement profile; (ii) register determining the extent of variation uncovered between the other lexical items. In both English and French, academic prose was found to display little lexical variation, whereas in the editorials, different CAs were found to display partly distinct placement preferences. This is consistent with the results obtained from the cross-register comparison of CA placement in English and French, where it was highlighted that register influences placement in very similar

ways in the two languages, with academic writing inducing thematic 1 uses of CAs, whereas the editorials displayed high frequencies of rhematic CAs. Therefore, although this chapter has underlined a number of marked differences in CA placement between English and French, the *mechanisms* underlying placement seem to work in very similar ways in the two languages.

8.6   Conclusion

This chapter has investigated and compared the placement patterns of English and French conjunctive adjuncts of contrast in two written registers, viz. academic writing and newspaper editorials. It has provided conclusive evidence of the influence not only of language, but also register and lexis on conjunctive adjunct placement. With respect to language, while the corpus study did not reveal any differences between English and French in terms of the possible positions available for conjunctive adjuncts, marked differences were uncovered with respect to the preferences of each language for certain positions within this common set of possibilities (or servitudes). Whereas rhematic 2 CAs were strikingly more frequent in French than in English, all the other positions – and, especially, the thematic 1 and the rhematic 1 slots – were more common in English than in French. These differences were observed in both registers, suggesting that the positional preferences of each language system play a greater role on CA placement than the communicative situation in which the markers are used – as was indeed confirmed by the application of the CART statistical method on the data. This result stood in contrast with the findings of previous English-French contrastive studies, which had reported that the kind of cross-linguistic differences uncovered between the two languages were partly dependent on the type of texts analysed. I thus suggested that the respective part played by language and register may depend on the linguistic phenomenon investigated.

While register did not appear to have an influence on the cross-linguistic differences between English and French, the results did reveal a significant impact of register on CA placement within each language system: in both English and French, the editorials displayed a significantly higher frequency of CAs used rhematically than the research articles, which were more strongly associated with the thematic 1 position. This register difference was attributed to the discourse functions achieved by CAs used in certain positions: it was demonstrated that, whereas CAs used in thematic 1 position tend to perform a strictly linking function, expressing a relation of contrast between the two units that they relate, markers occurring in rhematic positions usually perform a range of rhetorical effects in addition to their basic linking function – such as laying focus on one specific part of the message, high-

lighting a topical shift in discourse, or isolating the part of the sentence over which the CA has its main scope. It was thus postulated that, whereas writers of research articles tend to use CAs mostly to express purely logical connections between the parts of their discourse – in line with the predominantly informational purposes of the register – the authors of editorials make a greater use of the rhetorical potential afforded by CA placement, which may serve both the reader-friendly and the highly persuasive communicative concerns of the register.

The corpus results also provided evidence of the influence of lexis on CA placement, as some amount of variability was uncovered between the different markers within each language system. More precisely, lexis and register were found to interact to influence placement. Firstly, the amount of lexical variability was found to differ markedly across registers: in both English and French, very little lexical variation in placement was found in the research articles, whereas in the editorials, different markers appeared to exhibit partly distinct placement preferences. Thus, the amount of lexical variation was found to depend on the communicative situation in which the CAs were used. In addition, the corpus analysis shed light on two main types of placement profiles for CAs of contrast: on the one hand, some CAs (e.g. *instead, though, plutôt*) were found to display relatively stable placement properties across registers, thus exhibiting a strong degree of lexical priming. Other markers (e.g. *however, on the other hand, cependant*) rather appeared to be – at least partly – 'stylistically-primed', in that their placement preferences were found to vary, sometimes markedly, between registers. In other words, the amount of register variation in placement was shown to be dependent on the lexical item investigated. It is also important to stress that, although marked differences in preferred placement patterns were uncovered between English and French, the influence of lexis and register on the position of CAs was found to operate in very similar ways in the two languages: in both English and French, (i) thematic 1 CAs were preferred in academic writing, whereas rhematic CAs were markedly more frequent in the editorials; (ii) lexis was found to interact with register, with the editorials displaying much more lexical variation in placement than academic writing. In addition, the distinction between lexically-primed and stylistically-primed CAs was found to be relevant for the description of CA placement in both language systems.

Finally, from a more theoretical point of view, this chapter has offered a very good example of the benefits that can be gained from the adoption of a dialectical approach between Systemic Functional Linguistics and corpus linguistics, as advocated by researchers such as Butler (2004) or Halliday and Matthiessen (2004: 36). Such an approach consists in attempting to complement the weaknesses of each framework with the strengths of the other. In practice, this means that the corpus data can be used to inform and refine the SFL theory, making sure

that it can faithfully account for the range of functions and meanings observed in authentic linguistic productions; conversely, the SFL theory should be used to give shape to the linguistic patterns emerging from the corpus data, which are sometimes so diverse that it can appear difficult to make sense of them. In this chapter, the corpus data was used to adapt and refine the SFL system of thematic structure, by making distinctions within the rhematic subpart of the message. Such distinctions were shown to be essential in order to provide a detailed account of CA placement across languages: on the one hand, failing to make distinctions within the Rheme would have led us to miss important differences between the two languages, since English and French were found to differ markedly in their frequency of use of the different slots available within the Rheme. On the other hand, the various positions identified within the Rheme also appeared not to be equivalent from a discourse point of view: instead, they were each associated with their own range of rhetorical effects. This is also something that would have gone unnoticed if the Rheme had been viewed as one single entity, as is typically the case in the SFL literature. Corpus linguistics in turn profited from the Systemic Functional framework, which offered the necessary theoretical tools to interpret (some of) the corpus results – and more particularly, the differences between registers – which could be accounted for by reference to the SFL systems of thematic structure (e.g. with rhematic 1 CAs emphasising topical shifts or thematic continuity in discourse) and information structure (e.g. with the rhematic 2 CAs signalling a boundary between Given and New information).

CHAPTER 9

General conclusion

The overarching objective of this study has been to investigate and compare the use of conjunctive markers of contrast in English and French written language, taking a predominantly syntactic perspective towards these linguistic items. In line with the corpus approach adopted in this book, the description of conjunctive markers systematically started from quantitative discussions of the frequencies of CMs and their various usage patterns in the two languages. However, the tendencies emerging from these frequency comparisons quickly made clear that, in order to fully understand the syntactic patterning of English and French conjunctive markers of contrast, it was imperative to complement these quantitative observations with more qualitative and fine-grained analyses of the corpus data – an undertaking which was greatly facilitated by the combination of the corpus approach with the Systemic Functional theory. In particular, the syntax of the CMs appeared to be tightly intertwined with discourse and lexical factors. This general conclusion aims to (i) summarise the main findings yielded by this study; (ii) take stock of its main contributions to the various fields of linguistic research in which it has been grounded (viz. corpus linguistics, Systemic Functional Linguistics, discourse analysis and contrastive linguistics); and (iii) identify a number of promising avenues for future research opened up by the present study.

9.1 Summary of the main findings

As was announced in the introduction to this book, the analyses carried out in the present study were guided by two central research questions, viz.:

i. Do English and French differ in (a) their overall frequency of use of conjunctive markers of contrast; and (b) the types of conjunctive markers that they tend to prefer to signal contrast between two clauses or sentences?
ii. Do English and French differ with respect to the positions that they allow for and/or prefer for conjunctive adjuncts of contrast?

This section summarises the major findings emerging from the analyses performed in this study. It is subdivided into two main parts, each presenting the major answers to one of the research questions listed above.

9.1.1 Frequency and patterns of use of English and French conjunctive markers of contrast

The first major research question which has been at the centre of this study pertained to the frequency and types of CMs of contrast used in English and French written language and was tackled in Chapter 7. This part of the analysis relied primarily on the Mult-Ed comparable corpus of newspaper editorials. Based on the non-empirical, largely introspection-based English-French contrastive literature, two hypotheses were formulated in relation to this research question. Firstly, following the widespread assertion that "the general tendency seems to be to mark cohesion in French in a more explicit way than in English, using more linguistic material" (Armstrong 2005: 196), it was hypothesised that French would display a significantly higher overall frequency of explicit markers of contrast than English. Regarding the preferred types of markers in each language, the claims found in the contrastive literature led me to expect that relations of contrast would tend to be expressed by means of coordinators in English, and subordinators as well as conjunctive adjuncts in French.

The first of these hypotheses was strongly rejected by the corpus results, since explicit markers of contrast were found to occur markedly more frequently in the English than in the French subpart of Mult-Ed – with a ratio of about 1.5 English CM per French marker of contrast. Likewise, the supposed preference of English for coordinators, and French for subordinators and CAs, remained largely unsubstantiated by the corpus evidence. Firstly, in terms of frequencies, all three types of markers – rather than just coordinators – were found to be significantly more common in English than in French. With respect to proportions of use, on the other hand, only the preference of French for subordinators was confirmed; conjunctive adjuncts were found to be more common in English than in French, whilst no significant difference in proportion of contrastive coordinators was uncovered between the two languages. Complementary analyses carried out in the LOCRA comparable corpus of research articles suggested that these contrastive results were partly generalisable across (written) registers, since conjunctive adjuncts of contrast were also found to be markedly more frequent in English than in French – although this frequency difference was noticeably less extensive in the academic than in the editorial register. Since the analyses carried out in this book only included one category of explicit conjunctive markers, viz. CMs of contrast, the corpus results

did not permit a general rejection of the claims found in the contrastive literature. However, they clearly demonstrated the sheer necessity of at least modulating these long-standing, introspection-based statements with respect to the type of logico-semantic relationship holding between the discourse segments.

In a second stage, the cross-linguistic analysis of English and French CM use moved beyond the broad categorisation of markers into the three classes of 'coordinators', 'subordinators' and 'conjunctive adjuncts' to also include more fine-grained comparisons of the types of syntactic structures (viz. finite, non-finite or verbless clauses; main, hypotactic, embedded or minor clauses) in which these CMs were included. This kind of syntactic approach to CM use is uncommon in research on discourse relations, which has predominantly concentrated on the semantic features of conjunctive markers. Although such a perspective may have seemed surprising at first sight, it yielded illuminating results, which demonstrated the great value of more 'syntactically-oriented' analyses of conjunctive markers. More specifically, it was shown that different syntactic uses of conjunctive markers of contrast are associated with distinct rhetorical and/or stylistic effects at the discourse level. In other words, choices pertaining to the types of clauses hosting English and French conjunctive markers appeared to be situated at the interface between syntax and discourse. On the one hand, a number of syntactic uses of CMs (viz. markers used in minor clauses, sentence-initial uses of coordinators) were found to function as devices of syntactic fragmentation, and had the effect of emphasising not only the segment introduced by the marker, but also the discourse relation of contrast itself. Such patterns were shown to be in accordance with the increasingly informal and speech-like character of newspaper language, while also serving the highly persuasive communicative goals of the editorial register – by considerably enhancing the incisive and 'punchy' character of the writer's argumentation. Other syntactic patterns (e.g. CMs occurring in non-finite, verbless and/or hypotactic clauses) were shown to testify to some degree of syntactic compression in the expression of contrast, which appeared to go hand in hand with the space-saving concerns of journalistic writing. Accordingly, the differences in syntactic patterning of English and French CMs uncovered in the corpus data were demonstrated to reflect subtle divergences in the strategies of textual development adopted by the two languages with respect to the expression of contrast. The syntactic analysis of conjunctive markers also highlighted the impact of lexical factors on the patterning of CMs of contrast in both English and French, by uncovering marked differences in the preferred types of syntactic structures of CMs belonging to the same grammatical category. All in all, the use of English and French conjunctive markers of contrast was thus demonstrated to lie at the interface between syntax, discourse and lexis.

9.1.2   Placement patterns of English and French conjunctive
adjuncts of contrast

The second central objective of this research was to investigate and compare the
placement patterns of English and French conjunctive adjuncts of contrast. This
part of the study was based on both the Mult-Ed corpus of newspaper editorials,
and the LOCRA corpus of research articles, with a view to assessing the impact of
register variation on CA placement, both within and across languages. In addition,
the corpus analysis also aimed to evaluate the role played by lexis on the position of
English and French CAs of contrast. Two main hypotheses were tested in this part of
the study. Firstly, as French has sometimes been claimed to exhibit a greater degree
of syntactic flexibility than English (notably with respect to the canonical 'Subject
– Verb – Object' structure of the sentence), it was expected that French might offer
a larger variety of placement possibilities for conjunctive adjuncts as compared to
English. Secondly, based on the descriptions of CA placement found in the English
and the French monolingual literature, it was hypothesised that sentence-medial
CAs would be more frequent in French than in English, where a large majority of
the markers would appear at the beginning of the T-Unit. The first of these hypoth-
eses was rejected by the corpus results. No differences were found between the two
languages in terms of the possible positions that they make available for conjunctive
adjuncts of contrast. By contrast, the results pointed to clear differences with respect
to the positional preferences of English and French. However, these only partly con-
firmed the second research hypothesis. While sentence-initial (or 'thematic 1') CAs
were indeed found to be more typical of English than French, the medial positions
did not appear to be strictly specific to French. Rather, the two languages were found
to prefer different medial (or rhematic) positions for conjunctive adjuncts of con-
trast: whereas the rhematic 2 position (encompassing all the CAs used within the
predicate of the T-Unit) was markedly more frequent in French than in English,
the opposite was true of the rhematic 1 category, comprised of the CAs occurring
between the topical Theme and the main verb phrase of the T-Unit. Importantly –
and in the same way as the cross-linguistic differences in overall frequencies of con-
junctive adjuncts of contrast, cf. above – these frequency differences were found to
be stable across the academic and the editorial registers, suggesting that the respec-
tive preferences of the English and the French language systems influence CA place-
ment to a greater extent than the communicative situation in which these markers
are used. Such a conclusion was supported by the application of the multifactorial
statistical method of Classification and Regression Trees on the data, and contrasted
with previous contrastive results on the English-French language pair, which had
suggested that the type of cross-linguistic differences uncovered between the two
languages were partly dependent on the register analysed.

Even though register did not appear to influence the type of cross-linguistic differences uncovered between English and French, the corpus results did demonstrate a significant impact of register on CA placement *within* each language system: in both English and French, the editorials displayed a significantly higher proportion of rhematic CAs than the research articles, which were more closely associated with the thematic 1 position. One explanation put forward to account for such cross-register differences pertained to the rhetorical effects that may be achieved by CA placement. It was shown that, similarly to the syntactic patterning of conjunctive markers of contrast, CA placement could be said to be situated at the syntax-discourse interface, since different positional choices proved to be associated with distinct effects at the discourse level. More specifically, whereas thematic 1 (or sentence-initial) CAs appeared to perform a strictly connective function, the CAs used in all the other positions were shown to assume further functions in addition to their basic linking role – including laying focus on some part of the T-Unit, emphasising (breaks in) the thematic continuity of discourse, or clearly isolating the part of the message over which the contrastive CA has its main scope. Accordingly, the differences in ratio of thematic vs rhematic CAs in the academic and the editorial subcorpora were ascribed to the diverging communicative purposes of the two registers: as a predominantly informational text type, academic prose tends to use CAs mostly to express purely logical links between discourse units. By contrast, the type of rhetorical effects afforded by CAs placed rhematically were shown to be in line with both the persuasive and reader-friendly concerns of the editorial register.

Finally, the corpus analysis also identified lexis as an influential factor with respect to CA placement, with different conjunctive adjuncts displaying partly distinct positional preferences within each language system. In fact, in both English and French, lexis was found to interact with register to influence CA placement. On the one hand, the amount of lexical variation differed significantly between registers, with the placement patterns of English and French CAs being more diversified in the editorials than in the research articles. On the other hand, the amount of influence exerted by register on the placement of CAs differed across lexical items: while some markers displayed relatively stable placement preferences across registers – thus being strongly 'lexically-primed' – other markers were characterised by some degree of 'stylistic priming', in that they exhibited clearly distinct placement profiles in the editorials and in the research articles.

All in all, from the general summary of the main results obtained in the present study, it appears clearly that both types of syntactic analysis performed in this book (viz. Chapter 7 and Chapter 8) shed light on the existence of a tight connection between (i) the syntax of conjunctive markers, and (ii) the influence of discourse and lexical factors. In both parts of the study, the syntactic choices made with respect to conjunctive markers appeared to be motivated by (i) the (partly

unconscious) wish of the writer to create specific stylistic or rhetorical effects at discourse level; and (ii) the syntactic preferences of individual lexical items within each language system.

9.2   Main contributions of the study

This study was situated at the intersection between four main fields of linguistic research, viz. (i) corpus linguistics; (ii) Systemic Functional Linguistics; (iii) contrastive linguistics; and (iv) discourse analysis. This section underlines some of the main contributions that this book has made to these fields.

9.2.1   Contribution to (contrastive) discourse analysis

An overwhelming majority of the studies on discourse relations – whether mono- or multilingual – have focused predominantly on the semantic features of conjunctive markers. In comparison, the more syntactic aspects of these linguistic items have been largely overlooked in the discourse research to date. An excellent example of this imbalance pertains to the many attempts at unifying frameworks of annotation of discourse relations which have been made in recent years: while a great number of studies have been concerned with establishing common frameworks for the semantic annotation of discourse relations (see e.g. all the work carried out in the framework of the TextLink COST Action, led by Prof. L. Degand between 2014 and 2018), discourse annotation schemes still vary widely with respect to issues of discourse segmentation, i.e. in the criteria used to determine what linguistic segments qualify as discourse units – when such criteria are in fact made explicit (see also Sanders et al. 2018: 54–55 on this). Discrepancies as regards segmentation inevitably impede the comparability of the results, as studies that differ in the units which they consider as discourse segments necessarily differ in the number and type of conjunctive markers identified in the data. In addition, a lack of explicitness as to the segmentation principles applied makes it nearly impossible to replicate a given study on a different data set. One important contribution of this research has been to work towards more transparency in the segmentation of discourse, by putting forward very clear and explicit criteria to distinguish discourse (viz. clausal) from non-discourse (viz. non-clausal) segments. The study raised a number of important syntactic issues which arise when identifying explicit markers of coherence relations in authentic corpus data, and which had so far tended to be glossed over in discourse research.

Particular attention was granted to issues of cross-linguistic comparability with respect to discourse segmentation: I showed that the criteria that can be used

to identify a discourse segment in one language (e.g. the presence of a verb) are not necessarily equally appropriate for the analysis of other languages, and I demonstrated the necessity of ensuring the cross-linguistic neutrality of the segmentation criteria when approaching conjunctive markers from a contrastive perspective, so as not to bias the analysis in favour of one of the languages studied. More generally, this book has offered fairly deep theoretical reflections on various questions of cross-linguistic comparability for the contrastive analysis of discourse relations, by seeking to identify the best possible *tertium comparationis* at each level of the analysis. In addition to suggesting cross-linguistically valid principles of discourse segmentation, the present study also took special care to adopt a reliable *tertium comparationis* for cohesion and conjunctive markers, and demonstrated that the most appropriate basis for comparison of markers of cohesive conjunction across languages was a functional one: I decided to include English and French coordinators, subordinators *and* conjunctive adjuncts into the analysis, on the grounds that they all shared the function of expressing logico-semantic relations (such as causality, contrast or addition) explicitly between two clauses, sentences or paragraphs. Another important contribution of the study was to provide a thoroughly-circumscribed definition of contrast that was adapted to the description of both English and French conjunctive markers, and which was used as semantic *tertium comparationis* for the study. This was a challenging task, as the definitions of contrast available in the literature – whether monolingual, multilingual or language-neutral – were extremely diverse and usually very vague. After providing a comprehensive overview of the literature on contrast, I identified a number of central semantic components of the category, and defined contrast rather broadly as a class subsuming a set of three main relations (viz. opposition, concession and correction) which have in common that they signal a relation of mutual exclusiveness between two discourse segments within the same semantic domain. Again, this tripartite view of contrast was largely guided by concerns of cross-linguistic validity: on the one hand, it was the one which was obtained when merging the basic dichotomies made in the English and the French literature with respect to contrast; on the other hand, it was the definition that prevailed in most studies attempting to provide universally-valid descriptions of contrastive relations. In summary, this study has attempted to put forward a reliable theoretical framework for the contrastive study of conjunctive markers of contrast. Although strictly speaking, the cross-linguistic reflections presented in this book pertained to English and French, most of them are largely transposable – perhaps with minor adaptations – to the comparison of conjunctive marker use in other language pairs.

This study has also provided extensive evidence of the considerable relevance of approaching the analysis of conjunctive markers from the angle of syntax, rather than investigating these linguistic items mainly through the lens of semantics, as is

typically the case in current discourse research. The results emerging from the corpus analyses convincingly demonstrated that the syntactic choices associated with conjunctive markers are meaningful: they have true rhetorical potential, in that they may be used to create a range of stylistic or discourse effects, in line with the communicative goals of the registers in which the markers are used. Alongside stylistic factors, syntactic choices in terms of CM use were also shown to be influenced by lexical and language factors, since significant differences in syntactic patterning were uncovered between (i) languages; and (ii) various CMs within each language system. Such results were shown to have a number of implications for language learners and translators, for example. In view of the richness of the results emerging from these analyses, I believe that a syntactic approach to CMs opens a promising line of inquiry into CM use, by providing findings of an essentially new type.

9.2.2   Systemic Functional Linguistics and corpus linguistics

Another major contribution of this study has been to work towards a greater rapprochement between Systemic Functional Linguistics and corpus linguistics. As advocated by renowned SFL researchers such as Halliday and Matthiessen (2004: 36), Butler (2004) or Thompson and Hunston (2006a; 2006b), this study adopted a dialectical approach to SFL and corpus linguistics, in which the strengths of each framework were used to overcome the weaknesses of the other framework, in an attempt to make each approach benefit from the other. More specifically, the corpus data was used to refine the SFL theory, ensuring that it could faithfully account for the variety of usage patterns observed in authentic corpus data. Conversely, the SFL theory was used to interpret and give shape to the general tendencies emerging from the corpora, by relating them to a general conception or theory of language.

In this book, the main aspect of the SFL theory which was refined through the use of corpus methods was the textual metafunction, and more particularly the system of thematic structure. Based on the analysis of large amounts of authentic corpus data in English and French, it quickly became clear that the current SFL description of thematic structure, where all the linguistic content that follows the topical Theme is viewed as a single, undifferentiated entity, was not detailed enough to allow for a comprehensive description of the positional patterns of English and French conjunctive adjuncts of contrast. Amongst other problems, grouping all the conjunctive adjuncts occurring after the topical Theme within a single category would have caused me to overlook one central difference between English and French, viz. the preference of English for contrastive CAs occurring right after the topical Theme, as opposed to the propensity of French to use these CAs within the predicate of the T-Unit. The analysis of the corpus data thus prompted me to put forward further divisions within the rhematic part of the message, by identifying three rhematic positions for CAs, in addition to the two

thematic slots already available in the SFL theory. In this sense, the present study may be said to contribute to the "ambitious research programme" envisaged by Butler (2004: 176) over fifteen years ago:

> [W]e need an ambitious research programme in which corpora are used to tackle head-on the task of investigating which aspects of [the theory], including the most fundamental, we genuinely need to abandon in the face of the new evidence, which of them can remain, but with perhaps quite extensive modifications, and which are still robust enough to withstand the onslaught with only minor changes.

In this particular case, the analysis of corpus data neither confirmed, nor firmly rejected the SFL account of thematic structure – which respectively correspond to the third and the first possible outcomes of a combined SFL and corpus approach in Butler's quote. Rather, it pointed to the necessity of modulating or refining the existing system to accommodate the diversity of patterns observed in authentic language use (viz. the second of Butler's suggested outcomes). In addition, the analysis of a large, stylistically-varied bilingual corpus also allowed me to qualify the very tight connection made by SFL between conjunctive adjuncts and the thematic part of the message: the corpus data provided substantial evidence that the strength of this relationship depends on various factors, among which language (since thematic CAs were less common in French than in English) and register (with the editorials displaying a fairly large proportion of CAs used rhematically).

Conversely, the Systemic Functional framework greatly benefited the corpus analysis, by providing a range of theoretical constructs to help understand some of the main quantitative trends emerging from the data. For example, the extensive frequency differences observed between registers – viz. the marked preference of editorials for rhematic CAs as compared to the research articles – could be accounted for by reference to the SFL notions of Theme (and thematic progression), Rheme, Given and New. This made it possible to relate the corpus findings to a higher-order linguistic theory, thus avoiding the pitfall of providing isolated or theory-thin descriptions of individual areas of language, approached from an exclusively or predominantly quantitative perspective (see also next section on this issue). All in all, the integration of the corpus and the SFL frameworks in the present study thus provides an excellent illustration of the complementarity between the two frameworks of linguistic analysis, and of the considerable benefits that such a combined approach can offer.

9.2.3 Quantitative vs qualitative/macro vs micro linguistic research

As underlined by Gilquin (2010: 281), much research in corpus linguistics has been criticised for its penchant for 'number crunching', viz. its tendency to be

"exclusively concerned with the presentation of frequency data as a goal without couching the data within the framework of meaningful questions about the structure or usage of the language being studied" (Aarts 2000: 7). As highlighted by Aarts (ibid.), that kind of linguistic work "invariably elicits a 'so what' response", such as the following: "so what if thematic conjunctive adjuncts are significantly more frequent in academic prose than in newspaper editorials" or "so what if French conjunctive markers are more commonly found in minor clauses than their English counterparts?" (see also Granger 1998). One key contribution of this research has been to provide convincing evidence that the adoption of a large-scale, quantitative approach to linguistic analysis in no way precludes the inclusion of thorough qualitative reflections. The analyses carried out in this study were based on a corpus of over 8 million words in total, corresponding to 24,000 instances of conjunctive markers of contrast. Despite the fact that all these CMs were the object of general quantitative comparisons, supported by the use of statistical methods, the book also offered a wide variety of detailed qualitative discussions of the phenomena under investigation (e.g. comments on the discourse effects produced by distinct syntactic uses or positions of English and French conjunctive markers).

In fact, I believe that this study has demonstrated how complementary quantitative and qualitative linguistic analyses are. Quantitative observations constitute an excellent starting point to identify the elements that appear to be worthy of qualitative attention. The problem with 'purely qualitative' research, i.e. research which focuses on a small set of linguistic occurrences which are analysed in great detail, is that it tends to give rise to rather anecdotal observations. When qualitative comments are completely unguided by quantitative information, it is very difficult to identify what parts of the discussion are really significant for a linguistic description aiming at generalisability. Conversely, as pointed out above, purely quantitative linguistic research usually runs the risk of suffering from the 'so what' syndrome. A combined quantitative and qualitative perspective on linguistic description, by contrast, ensures both that (i) the quantitative tendencies observed in the data are – at least partly – *explained*, rather than being simply stated, with no further specification of the ways in which they contribute to the development of linguistic knowledge; and (ii) the linguistic phenomena which are the object of thorough – and usually quite time-consuming – qualitative analyses are truly worthy of discussion, in that they can help explain some significant frequency patterns observed in the corpus data.

In addition – and in connection with what precedes, this research showed very clearly how analyses at the macro level do not prevent the study of phenomena at a more fine-grained (or micro) level of the linguistic system: although the present study took as its starting point a rather broad, onomasiologically-defined category of linguistic items (viz. CMs of contrast in English and French), analysed

in a very large data set, it also included a range of fairly focused analyses (e.g. analyses pertaining to the idiosyncratic syntactic patterning or placement patterns of individual conjunctive markers within each language system). In fact, as this study showed, not only do macro analyses *allow for* the study of more fine-grained linguistic phenomena, but the combination of a macro perspective towards linguistic research with more micro analyses of the phenomena under scrutiny also makes it possible to put some of the corpus findings in perspective: for example, it is only by starting from the general category of conjunctive markers of contrast that the syntactic idiosyncrasies of some of the markers could emerge from the analysis. While case studies focusing on one or two markers make it possible to provide very detailed descriptions of the items under investigation, they usually do not allow the researcher to situate these findings in a broader framework, by identifying in what ways the items analysed are representative of the category to which they belong, or rather display specific usage features.

9.3 Promising avenues for future research

This book has offered numerous new insights into the use of conjunctive markers of contrast in English and French. However, the study does not claim to have exhausted the research possibilities relating to this object of study. This section intends to identify some remaining research avenues with respect to English and French CMs of contrast. One such promising line of inquiry is semantics. As explained above, the main point of departure in this research has been the syntactic features of conjunctive markers of contrast. I have shown how fruitful a more syntactically-oriented approach to CM use can be. However, the decision to focus on the syntax of conjunctive markers has been made at the expense of fine-grained semantic analyses. For example, while great care was taken to provide a well-documented, cross-linguistically valid definition of contrast, and to meticulously select the markers in the corpus whose meanings corresponded to that definition, I did not systematically identify the subtype of contrast (viz. opposition, concession and correction) expressed by each CM token in the data set. This was also largely due to the sheer volume of data analysed here: in view of the great time and effort required by the mere semantic disambiguation of the data (viz. distinguishing between contrastive and non-contrastive uses of the markers automatically extracted from the corpus), I quickly realised that coding each CM token semantically would not be feasible in the framework of the present study, which necessitated a certain 'trade-off' between the quantity of data analysed, and the precision of the semantic analyses.

However, such fine-grained semantic inquiries might provide additional insights into the differences between English and French. For instance, although

relations of contrast in general are supposedly less prone to implicitation than other logico-semantic categories, some studies have demonstrated that there is also variation across the subtypes of contrast in the extent to which they need to be signalled explicitly by means of a CM. Asr and Demberg (2012), for example, have shown that relations of concession are particularly unlikely to be left implicit, due to the high degree of cognitive complexity that they entail. In a future study, it would be worth investigating whether the gap in overall frequency of CMs of contrast observed between English and French mostly pertains to relations of opposition and correction, or whether all three subtypes of contrast are expressed more explicitly in English than in French. In addition to contributing to the interpretation of the differences in overall frequencies, coding the data semantically could also help to better understand some of the frequency patterns of individual CMs identified in Chapter 7, both across languages (see e.g. the frequency difference between *cependant* and *however*) and registers (see e.g. the predominance of *yet/pourtant* in the editorials, as opposed to *however/cependant* in the research articles). Yet another aspect of CM use which could benefit from more detailed semantic analyses is conjunctive adjunct placement. For example, in Chapter 8, it was shown that the rhematic 1 position was especially well-suited for relations of opposition, which involve a double contrast between the segments related by the marker. In a future study, it might be worth coding the data for both position *and* type of contrast to identify possible correlations between the positional choices made by writers and the subtype of contrast expressed by the markers. Finally, it would also be extremely useful to replicate the present study on conjunctive markers expressing other types of logico-semantic relationships than contrast, so as to establish whether the significantly greater degree of cohesive explicitness of English as compared to French is (partly) specific to contrast, or can be generalised to other discourse relations.

Secondly, complementary analyses of conjunctive markers in translation data – in line with the combined approach advocated by Johansson (2007) – appear to be a logical follow-up to the comparable corpus analyses presented here. For instance, in view of the striking gap in frequency of markers of contrast uncovered between the two languages, a study of CM omissions and, especially additions, in English-French translation data would help identify the types of devices which French uses in contexts where English resorts to an explicit marker of contrast (e.g. mere juxtaposition, logical metaphor, special syntactic structures such as thematisation, stance markers such as *certes, malheureusement*, etc.). In addition, I would personally be very interested in complementing the study of conjunctive adjunct placement carried out on the basis of comparable data with the analysis of translated texts, with a view to assessing the translators' degree of awareness of the inherent word order preferences of their target language. Such a study would

be especially interesting since, as highlighted by Loock (2018), issues pertaining to word order are one of the areas where human translators still have a true edge over machine translation, making it all the more important for them to master those aspects of the translation practice. It could also be extremely interesting to investigate how translators handle the range of rhetorical functions that may be performed by CA placement. In this respect, one might expect translators to display a tendency towards normalisation: it could be hypothesised that many CAs used rhematically in the source texts will tend to be moved to thematic 1 position – which constitutes the most rhetorically 'neutral' or 'unmarked' slot for these items – in the target texts, in line with Baker's (1996: 184) statement that marked syntactic structures are often normalised in translation. This is, in any case, the pattern that was identified in learner corpus research, where a number of studies have underlined an overuse of sentence-initial conjunctive adjuncts by learners of English (cf. e.g. Field & Yip 1992; Granger & Tyson 1996).

Finally, this book concentrated on three main factors of influence with respect to CM use, viz. language, lexis and register. It is clear, however, that other factors of influence may also be worthy of investigation. With respect to conjunctive adjunct placement, for instance, one aspect that would certainly warrant further attention is the linguistic environment of the markers (e.g. the complexity of the verb phrase, the length of the subject/object, the presence of a fronted subclause or another conjunctive marker in initial position, etc.). I very much look forward to assessing the impact of these additional factors in my future research.

References

Aarts, Bas. 2000. Corpus linguistics, Chomsky and fuzzy tree fragments. In *Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20)*, Christian Mair & Marianne Hundt (eds), 5–14. Amsterdam: Rodopi.

Adam, Jean-Michel. 2008. *La linguistique textuelle: Introduction à l'analyse textuelle des discours*. Paris: Armand Colin.

Aijmer, Karin. 2008. Comparable and parallel corpora. In *Corpus Linguistics: An International Handbook*, Vol. 1, Anke Lüdeling & Merja Kytö (eds), 275–292. Berlin: Walter de Gruyter.

Aijmer, Karin & Altenberg, Bengt. 1996. Introduction. In *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies, Lund 4–5 March 1994*, Karin Aijmer, Bengt Altenberg & Mats Johansson (eds), 11–16. Lund: Lund University Press.

Aijmer, Karin & Altenberg, Bengt. 2000. The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies. In *Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20)*, Christian Mair & Marianne Hundt (eds), 15–33. Amsterdam: Rodopi.

Aijmer, Karin, Altenberg, Bengt & Johansson, Mats. 1996. Text-based contrastive studies in English. Presentation of a project. In *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies, Lund 4–5 March 1994*, Karin Aijmer, Bengt Altenberg & Mats Johansson (eds), 73–85. Lund: Lund University Press.

Aijmer, Karin & Lewis, Diana. 2017. *Contrastive Analysis of Discourse-pragmatic Aspects of Linguistic Genres*. Berlin: Springer. https://doi.org/10.1007/978-3-319-54556-1

Aijmer, Karin & Simon-Vandenbergen, Anne-Marie. 2011. Pragmatic markers. In *Discursive Pragmatics* [Handbook of Pragmatics Highlights 8], Jan Zienkowski, Jan-Ola Östman & Jef Verschueren (eds), 223–247. Amsterdam: John Benjamins. https://doi.org/10.1075/hoph.8.13aij

Alonso Belmonte, Isabel. 2007. Newspaper editorials and comment articles: a "Cinderella" genre? *RAEL: Revista Electrónica de Lingüística Aplicada* 1: 1–9.

Altenberg, Bengt. 1984. Causal linking in spoken and written English. *Studia Linguistica* 38(1): 20–69. https://doi.org/10.1111/j.1467-9582.1984.tb00734.x

Altenberg, Bengt. 1986. Contrastive linking in spoken and written English. In *English in Speech and Writing. A Symposium*, Gunnel Tottie & Ingegerd Bäckland (eds), 13–40. Uppsala: Acta Universitatis Upsaliensis.

Altenberg, Bengt. 1998. Connectors and sentence openings in English and Swedish. In *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, Stig Johansson & Signe Oksefjell (eds), 115–143. Amsterdam: Rodopi.

Altenberg, Bengt. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In *Out of Corpora. Studies in Honour of Stig Johansson*, Hilde Hasselgård & Signe Oksefjell (eds), 249–268. Amsterdam: Rodopi.

Altenberg, Bengt. 2002. Concessive connectors in English and Swedish. In *Information Structure in a Cross-linguistic Perspective*, Hilde Hasselgård, Stig Johansson, Bergljot Behrens & Cathrine Fabricius-Hansen (eds), 21–43. Amsterdam: Rodopi. https://doi.org/10.1163/9789004334250_003

Altenberg, Bengt. 2006. The function of adverbial connectors in second initial position in English and Swedish. In *Pragmatic Markers in Contrast*, Karin Aijmer & Anne-Marie Simon-Vandenbergen (eds), 11–37. Oxford: Elsevier.

Altenberg, Bengt. 2007. The correspondence of resultive connectors in English and Swedish. *NJES: Nordic Journal of English Studies* 6(1). <http://hdl.handle.net/2077/4521> (4 September 2020).

Altenberg, Bengt & Aijmer, Karin. 2002. Zero translations and cross-linguistic equivalence: Evidence from the English-Swedish Parallel Corpus. In *From the COLT's Mouth … and Others'. Language Corpora Studies in Honour of Anna-Brita Stenström*, Leiv Egil Breivik & Angela Hasselgren (eds), 19–41. Amsterdam: Rodopi.

Altenberg, Bengt & Granger, Sylviane. 2002. Recent trends in cross-linguistic lexical studies. In *Lexis in Contrast: Corpus-based Approaches* [Studies in Corpus Linguistics 7], Bengt Altenberg & Sylviane Granger (eds), 3–48. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.7.04alt

Altenberg, Bengt & Tapper, Marie. 1998. The use of adverbial connectors in advanced Swedish learner's written English. In *Learner English on Computer*, Sylviane Granger (ed.), 80–93. London: Longman.

Ansary, Hasan & Babaii, Esmat. 2005. The generic integrity of newspaper editorials: A systemic functional perspective. *RELC Journal* 36(3): 271–295. https://doi.org/10.1177/0033688205060051

Anscombre, Jean-Claude & Ducrot, Oswald. 1977. Deux "mais" en français? *Lingua* 43: 23–40. https://doi.org/10.1016/0024-3841(77)90046-8

Armstrong, Nigel. 2005. *Translation, Linguistics, Culture: A French-English Handbook*. Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853598074

Asher, Nicholas & Lascarides, Alex. 2003. *Logics of Conversation*. Cambridge: CUP.

Asr, Fatemeh & Demberg, Vera. 2012. Implicitness of discourse relations. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, 2669–2684.

Astington, Eric. 1983. *Equivalences*. Cambridge: CUP.

Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In *Text and Technology*, Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds), 233–250. Amsterdam: John Benjamins. https://doi.org/10.1075/z.64.15bak

Baker, Mona. 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target* 7(2): 223–243. https://doi.org/10.1075/target.7.2.03bak

Baker, Mona. 1996. Corpus-based translation studies. The challenges that lie ahead. In *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager* [Benjamins Translation Library, 18] Harold Somers (ed.), 175–186. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.18.17bak

Balažic Bulc, Tatjana & Gorjanc, Vojko. 2015. The position of connectors in Slovene and Croatian student academic writing: A corpus-based approach. In *Meaning Making in Text*, Sonja Starc, Carys Jones & Arianna Maiorani (eds), 51–71. Houndmills: Palgrave Macmillan. https://doi.org/10.1057/9781137477309_4

Ballard, Michel. 1995. La traduction de la conjonction "and" en français. In *Relations discursives et traduction*, Michel Ballard (ed.), 221–293. Villeneuve d'Ascq: Presses universitaires de Lille.

Banks, David. 2002. Systemic functional linguistics as a model for text analysis. *ASp. La revue du GERAS* 35–36: 23–34. https://doi.org/10.4000/asp.1584

Banks, David. 2017. *A Systemic Functional Grammar of French*. London: Routledge. https://doi.org/10.4324/9781315228327

Banks, David, Eason, Simon & Ormrod, Janet. 2009. *La linguistique systémique fonctionnelle et la langue française*. Paris: L'Harmattan.

Becher, Viktor. 2011. When and why do translators add connectives? A corpus-based study. *Target* 23(1): 26–47. https://doi.org/10.1075/target.23.1.02bec

Bednarek, Monika. 2010. Corpus linguistics and systemic functional linguistics: Interpersonal meaning, identity and bonding in popular culture. In *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation*, Monika Bednarek & James R. Martin (eds), 237–266. London: Continuum.

Beeching, Kate. 2013. A parallel corpus approach to investigating semantic change. In *Advances in Corpus-based Contrastive Linguistics: Studies in honour of Stig Johansson* [Studies in Corpus Linguistics 54], Karin Aijmer & Bengt Altenberg (eds), 103–126. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.54.07bee

Bell, David. 2007. Sentence-initial "and" and "but" in academic writing. *Pragmatics* 17(2): 183–201. https://doi.org/10.1075/prag.17.2.01bel

Benamara, Farah & Taboada, Maite. 2015. Mapping different rhetorical relation annotations: A proposal. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, 147–152. Denver CO: Association for Computational Linguistics. https://doi.org/10.18653/v1/S15-1016

Berk, Richard A. 2016. *Statistical Learning from a Regression Perspective*. New York NY: Springer. https://doi.org/10.1007/978-3-319-44048-4

Berry, Margaret. 1996. What is theme? A(nother) personal view. In *Meaning and Form: Systemic Functional Interpretations*, Margaret Berry, Christopher Butler, Robin Fawcett & Guowen Huang (eds), 1–64. Norwood NJ: Ablex.

Berry, Margaret, Thompson, Geoff & Hillier, Hilary. 2014. Theme and variations. In *Theory and Practice in Functional-Cognitive Space* [Studies in Functional and Structural Linguistics 68], María de los Ángeles Gómez González, Francisco José Ruiz de Mendoza Ibáñez & Francisco Gonzálvez García (eds), 107–126. Amsterdam: John Benjamins.

Bestgen, Yves. 2017. Getting rid of the Chi-square and Log-likelihood tests for analysing vocabulary differences between corpora. *Quaderns de Filologia: Estudis Lingüístics* 22: 33–36. https://doi.org/10.7203/qf.22.11299

Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511621024

Biber, Douglas. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511519871

Biber, Douglas. 2004. Compressed noun-phrase structures in newspaper discourse: The competing demands of popularization vs. economy. In *New Media Language*, Jean Aitchison & Diana Lewis (eds), 169–181. London: Routledge.

Biber, Douglas & Conrad, Susan. 2001. Register variation: A corpus approach. In *The Handbook of Discourse Analysis*, Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton (eds), 175–196. Oxford: Blackwell.

Biber, Douglas & Gray, Bethany. 2010. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9(1): 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.

Bilger, Mireille & Cappeau, Paul. 2013. Une conjonction qui subordonne rarement: Le cas de "alors que." In *Morphologie, syntaxe et sémantique des subordonnants*, Colette Bodelot (ed.), 259–273. Clermont-Ferrand: Presses universitaires Blaise Pascal.

Blakemore, Diane. 1989. Denial and contrast: A relevance theoretic analysis of "but." *Linguistics and Philosophy* 12(1): 15–37. https://doi.org/10.1007/BF00627397

Blakemore, Diane. 1992. *Understanding Utterances: An Introduction to Pragmatics*. Oxford: Blackwell.

Blakemore, Diane. 2004. *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: CUP.

Bloor, Thomas & Bloor, Meriel. 2004. *The Functional Analysis of English: A Hallidayan Approach*. London: Arnold. https://doi.org/10.4324/9780203774854

Blum-Kulka, Shoshana. 1986. Shifts of cohesion and coherence in translation: Discourse and cognition in translation and second language acquisition studies. In *Interlingual and Intercultural Communication*, Juliane House & Shoshana Blum-Kulka (eds), 17–35. Tübingen: Gunter Narr.

Bolívar, Adriana. 2002. The structure of newspaper editorials. In *Advances in Written Text Analysis*, Malcolm Coulthard (ed.), 276–294. London: Routledge.

Bolton, Kingsley, Nelson, Gerald & Hung, Joseph. 2002. A corpus-based study of connectors in student writing. *International Journal of Corpus Linguistics* 7(2): 165–182. https://doi.org/10.1075/ijcl.7.2.02bol

Boularès, Michèle & Frérot, Jean-Louis. 2017. *Grammaire progressive du français. Niveau avancé*. Paris: Clé International.

Bourgoin, Charlotte. 2017. The role of the English it-cleft and the French c'est-cleft in research discourse. *Discours* (21): 3–31. https://doi.org/10.4000/discours.9366

Bourmayan, Anouch, Loiseau, Yves, Rimbert, Odile & Taillandier, Isabelle. 2017. *Grammaire essentielle du français. Niveau B2*. Paris: Didier.

Bowker, Lynne. 1999. Exploring the potential of corpora for raising language awareness in student translators. *Language Awareness* 8(3-4): 160–173. https://doi.org/10.1080/09658419908667126

Bowker, Lynne. 2003. Corpus-based applications for translator training: Exploring the possibilities. In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, Sylviane Granger, Jacques Lerot & Stephanie Petch-Tyson (eds), 169–183. Amsterdam: Rodopi.

Breiman, Leo, Friedman, Jerome, Olshen, Richard & Stone, Charles. 1984. *Classification and Regression Trees*. Belmont CA: Wadsworth.

Burchfield, Robert W. 2004. *Fowler's Modern English Usage*. Oxford: OUP.

Butler, Christopher. 2003. *Structure and Function: A Guide to Three Major Structural-Functional Theories: Part 2: From Clause to Discourse and Beyond* [Studies in Language Companion Series 64]. Amsterdam: John Benjamins.

Butler, Christopher S. 2004. Corpus studies and functional linguistic theories. *Functions of Language* 11(2): 147–186. https://doi.org/10.1075/fol.11.2.02but

Butt, David, Fahey, Rhondda & Feez, Susan. 2000. *Using Functional Grammar. An Explorer's Guide*. Sydney: National Centre for English Language Teaching and Research, Macquarie University.

Buysse, Lieven. 2017. English "so" and Dutch "dus" in a parallel corpus: An investigation into their mutual translatability. In *Contrastive Analysis of Discourse-pragmatic Aspects of Linguistic Genres*, Karin Aijmer & Diana Lewis (eds), 33–61. Berlin: Springer. https://doi.org/10.1007/978-3-319-54556-1_3

Caffarel, Alice. 2006. *A Systemic Functional Grammar of French: From Grammar to Discourse*. London: Continuum.

Caffarel, Alice, Martin, James R. & Matthiessen, Christian M. I. M. 2004. *Language Typology: A Functional Perspective* [Current Issues in Linguistic Theory 253]. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.253

Caffarel-Cayron, Alice & Rechniewski, Elizabeth. 2014. Exploring the generic structure of French editorials from the perspective of systemic functional linguistics. *Journal of World Languages* 1(1): 18–37. https://doi.org/10.1080/21698252.2014.893672

Carlson, Lynn & Marcu, Daniel. 2001. *Discourse tagging reference manual*. ISI technical report ISI-TR-545.

Carter, Ronald & McCarthy, Michael. 2006. *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge: CUP.

Carter-Thomas, Shirley. 2002. Theme and information structure in French and English: A contrastive study of journalistic clefts. In *14th Euro-International Systemic Functional Workshop - EISFW 2002*, C. A. M. Gouveia, C. Lopes da Costa, E. Ribeiro Pedro, L. Azuaga & S. Barcelos (eds). Lisbon: University of Lisbon.

Cartoni, Bruno, Zufferey, Sandrine & Meyer, Thomas. 2013. Annotating the meaning of discourse connectives by looking at their translation: The translation-spotting technique. *Dialogue & Discourse* 4(2): 65–86. https://doi.org/10.5087/d&d.v4i2.2823

Cartoni, Bruno, Zufferey, Sandrine, Meyer, Thomas & Popescu-Belis, Andrei. 2011. How comparable are parallel corpora? Measuring the distribution of general vocabulary and connectives. In *Proceedings of the 4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web (BUCC '11)*, 78–86. Stroudsburg PA: Association for Computational Linguistics.

Castagnoli, Sara. 2009. Regularities and Variations in Learner Translations: A Corpus-based Study of Conjunctive Explicitation. PhD dissertation, Università di Pisa.

Celce-Murcia, Marianne & Larsen-Freeman, Diane. 1999. *The Grammar Book: An ESL/EFL Teacher's Course*. Boston MA: Heinle & Heinle.

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In *Subject and Topic*, Charles Li (ed.), 25–55. New York NY: Academic Press.

Chafe, Wallace. 1982. Integration and involvement in speaking, writing, and oral literature. In *Spoken and Written Language: Exploring Orality and Literacy*, Deborah Tannen (ed.), 35–53. Norwood NJ: Ablex.

Chalker, Sylvia. 1996. *Linking Words*. London: Harper Collins.

Chang, Yu-Ying & Swales, John. 1999. Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers? In *Writing: Texts, Processes and Practices*, Christopher N. Candlin & Ken Hyland (eds), 145–167. London: Longman.

Charaudeau, Patrick. 1992. *Grammaire du sens et de l'expression*. Paris: Hachette.

Charolles, Michel & Vigier, Daniel. 2005. Les adverbiaux en position préverbale: Portée cadrative et organisation des discours. *Langue française* 148(4): 9–30. https://doi.org/10.3406/lfr.2005.6604

Chesterman, Andrew. 1998. *Contrastive Functional Analysis* [Pragmatics & Beyond New Series 47]. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.47

Christiansen, Thomas. 2011. *Cohesion: A Discourse Perspective*. Bern: Peter Lang. https://doi.org/10.3726/978-3-0351-0234-5

Chuquet, Hélène & Paillard, Michel. 1987. *Approche linguistique des problèmes de traduction anglais-français*. Paris: Ophrys.

Chuquet, Hélène & Paillard, Michel. 2017. *Glossaire de linguistique contrastive. Anglais-français*. Paris: Ophrys.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. New York NY: Academic Press.

Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112(1): 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Connor, Ulla & Moreno, Ana. 2005. Tertium comparationis: A vital component in contrastive research methodology. In *Directions in Applied Linguistics. Essays in Honour of Robert B. Kaplan*, Paul Bruthiaux, Dwight Atkinson, William Eggington, William Grabe & Vaidehi Ramanathan (eds), 153–164. Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853598500-015

Conrad, Susan. 1999. The importance of corpus-based research for language teachers. *System* 27(1): 1–18. https://doi.org/10.1016/S0346-251X(98)00046-3

Conrad, Susan. 2004. Corpus linguistics, language variation, and language teaching. In *How to Use Corpora in Language Teaching* [Studies Corpus Linguistics 12], John Sinclair (ed.), 67–85. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.12.08con

Conrad, Susan & Biber, Douglas. 2000. Adverbial marking of stance in speech and writing. In *Evaluation in Text: Authorial Stance and the Construction of Discourse*, Susan Hunston & Geoff Thompson (eds), 56–73. Oxford: OUP.

Cosme, Christelle. 2004. Towards a corpus-based cross-linguistic study of clause combining. Methodological framework and preliminary results. *BELL: Belgian Journal of English Language and Literatures* 2: 199–224.

Cosme, Christelle. 2006. Clause combining across languages: A corpus-based study of English-French translation shifts. *Languages in Contrast* 6(1): 71–108. https://doi.org/10.1075/lic.6.1.04cos

Cosme, Christelle. 2008a. A corpus-based perspective on clause linking patterns in English, French and Dutch. In *"Subordination" versus "Coordination" in Sentence and Text* [Studies in Language Companion Series, 98], Cathrine Fabricius-Hansen & Wiebke Ramm (eds), 89–114. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.98.06cos

Cosme, Christelle. 2008b. Clause Linking across Languages. A Corpus-based Study of Coordination and Subordination in English, French and Dutch. PhD dissertation, Université catholique de Louvain.

Cotter, Colleen. 2003. Prescription and practice. *Journal of Historical Pragmatics* 4(1): 45–74. https://doi.org/10.1075/jhp.4.1.04cot

Craggs, Richard & McGee Wood, Mary. 2005. Evaluating discourse and dialogue coding schemes. *Computational Linguistics* 31(3): 289–296. https://doi.org/10.1162/089120105774321109

Crewe, W. J. 1990. The illogic of logical connectives. *ELT Journal* 44(4): 316–325. https://doi.org/10.1093/elt/44.4.316

Crible, Ludivine. 2017. Discourse markers and (dis)fluency in English and French: Variation and combination in the DisFrEn corpus. *International Journal of Corpus Linguistics* 22(2): 242–269. https://doi.org/10.1075/ijcl.22.2.04cri

Crible, Ludivine. 2018. *Discourse Markers and (Dis)fluency: Forms and Functions across Languages and Registers* [Pragmatics & Beyond New Series 286]. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.286

Crible, Ludivine & Zufferey, Sandrine. 2015. Using a unified taxonomy to annotate discourse markers in speech and writing. In *Proceedings of the 11th Joint ACL - ISO Workshop on Interoperable Semantic Annotation (isa-11)*, Harry Bunt (ed.), 14–22. London: Association for Computational Linguistics.

Crismore, Avon, Markkanen, Raija & Steffensen, Margaret. 1993. Metadiscourse in persuasive writing. A study of texts written by American and Finnish university students. *Written Communication* 10(1): 39–71. https://doi.org/10.1177/0741088393010001002

Cross, Cate & Oppenheim, Charles. 2006. A genre analysis of scientific abstracts. *Journal of Documentation* 62(4): 428–446. https://doi.org/10.1108/00220410610700953

Csüry, István. 2001. *Le champ lexical de "mais": Étude lexico-grammaticale des termes d'opposition du français contemporain dans un cadre textologique*. Debrecen: Kossuth Egyetemi Kiadó.

Csüry, István. 2006. L'identification d'une discipline: Difficultés et perspectives. Commentaires sur les différentes approches d'un objet pluridisciplinaire. In *Des discours aux textes: Modèles et analyses*, Philippe Lane (ed.), 103–124. Mont-Saint-Aignan: Publications des Universités de Rouen et du Havre.

Cuenca, Maria Josep. 2003. Two ways to reformulate: A contrastive analysis of reformulation markers. *Journal of Pragmatics* 35(7): 1069–1093.
https://doi.org/10.1016/S0378-2166(03)00004-3

Cuenca, Maria Josep & Bach, Carme. 2007. Contrasting the form and use of reformulation markers. *Discourse Studies* 9(2): 149–175. https://doi.org/10.1177/1461445607075347

Cummings, Michael. 2009. The theme/rheme distinction and the method of development of written French text. In *La linguistique systémique fonctionnelle et la langue française*, David Banks, Simon Eason & Janet Ormrod (eds), 43–66. Paris: L'Harmattan.

Dafouz-Milne, Emma. 2008. The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics* 40(1): 95–113.
https://doi.org/10.1016/j.pragma.2007.10.003

Danlos, Laurence, Colinet, Margot & Steinlin, Jacques. 2015a. FDTB1, première étape du projet French Discourse Treebank: Repérage des connecteurs de discours en corpus. *Discours* 17. https://doi.org/10.4000/discours.9065

Danlos, Laurence, Colinet, Margot & Steinlin, Jacques. 2015b. *FDTB: Annotation des connecteurs de discours dans un corpus français*. Rapport technique. Paris: ALPAGE, Université Paris Diderot.

Danlos, Laurence & Roze, Charlotte. 2011. *Hiérarchie des relations de discours dans le FDTB*. Rapport technique. Paris: ALPAGE, Université Paris Diderot.

Das, Debopam & Taboada, Maite. 2013. Explicit and implicit coherence relations: A corpus study. In *Proceedings of the 2013 Annual Conference of the Canadian Linguistic Association*, Shan Luo (ed.). Victoria: University of Victoria.

De Beaugrande, Robert & Dressler, Wolfgang U. 1981. *Introduction to Text Linguistics*. London: Longman. https://doi.org/10.4324/9781315835839

de Carvalho, Paulo. 2005. Phrase nominale, 'parties du discours' et théorie syntaxique. *Syntaxe et Sémantique* 6: 87–102. https://doi.org/10.3917/ss.006.0087

De Cesare, Anna-Maria. 2013. On the focusing function of focusing adverbs: A discussion based on Italian data. *Linguistik Online* 44(4). https://doi.org/10.13092/lo.44.406

Degaetano-Ortlieb, Stefania. 2015. Evaluative meaning and cohesion: The structuring function of evaluative meaning in scientific writing. *Discours* 16: 3–29.
https://doi.org/10.4000/discours.9053

Degand, Liesbeth. 2004. Contrastive analyses, translation and speaker involvement: the case of "puisque" and "aangezien." In *Language, Culture and Mind*, Michel Achard & Suzanne Kemmer (eds), 251–270. Stanford CA: CSLI.

Degand, Liesbeth. 2014. "So very fast very fast then" Discourse markers at left and right periphery in spoken French. In *Discourse Functions at the Left and Right Periphery: Crosslinguistic Investigations of Language Use and Language Change*, Kate Beeching & Ulrich Detges (eds), 151–178. Leiden: Brill. https://doi.org/10.1163/9789004274822_008

Degand, Liesbeth & Fagard, Benjamin. 2011. "Alors" between discourse and grammar: The role of syntactic position. *Functions of Language* 18(1): 29–56.
https://doi.org/10.1075/fol.18.1.02deg

Degand, Liesbeth, Lefèvre, Nathalie & Bestgen, Yves. 1999. The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design* 1(1): 39–51.
https://doi.org/10.1075/dd.1.1.06deg

Degand, Liesbeth & Pander Maat, Henk. 2003. A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In *Usage-Based Approaches to Dutch. Lexicon, Grammar, Discourse*, Arie Verhagen & Jeroen van de Weijer (eds), 175–199. Utrecht: LOT.

Degand, Liesbeth & Sanders, Ted J. M. 2002. The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing* 15(7): 739–757.
https://doi.org/10.1023/A:1020932715838

Delaere, Isabelle & De Sutter, Gert. 2017. Variability of English loanword use in Belgian Dutch translations: Measuring the effect of source language, register, and editorial intervention. In *Empirical Translation Studies: New Methodological and Theoretical Traditions*, Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (eds), 81–112. Berlin: Mouton de Gruyter.
https://doi.org/10.1515/9783110459586-004

Delisle, Jean. 2013. *La traduction raisonnée. Manuel d'initiation à la traduction professionnelle de l'anglais vers le français*. Ottawa: University of Ottawa Press.

Demberg, Vera, Asr, Fatemeh & Scholman, Merel. 2017. How compatible are our discourse annotations? Insights from mapping RST-DT and PDTB annotations. *ArXiv e-prints*.
<http://arxiv.org/abs/1704.08893> (4 September 2020).

Dixon, Robert & Aikhenvald, Alexandra. 2011. *The Semantics of Clause Linking: A Cross-Linguistic Typology*. Oxford: OUP.

Downing, Angela. 1991. An alternative approach to theme: A systemic-functional perspective. *Word* 42(2): 119–143. https://doi.org/10.1080/00437956.1991.11435835

Dupont, Maïté. 2013. Contrastive Relations in English and French Editorials. A Corpus-based Study. MA dissertation, Université catholique de Louvain.

Dupont, Maïté. 2015. Word order in English and French: The position of English and French adverbial connectors of contrast. *English Text Construction* 8(1): 88–124.
https://doi.org/10.1075/etc.8.1.04dup

Dupont, Maïté. 2018. Between lexis and discourse. A cross-register study of connectors of contrast. In *Corpora and Lexis*, Sebastian Hoffmann, Andrea Sand, Sabine Arndt-Lappe & Lisa Marie Dillmann (eds.), 173–208. Leiden: Brill Rodopi.

Dupont, Maïté & Sandrine Zufferey. 2017. Methodological issues in the use of directional parallel corpora: A case study of English and French concessive connectives. *International Journal of Corpus Linguistics* 22(2): 270–297. https://doi.org/10.1075/ijcl.22.2.05dup

Ebeling, Jarle & Oksefjell Ebeling, Signe. 2013. *Patterns in Contrast* [Studies in Corpus Linguistics 58]. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.58

Eggins, Suzanne. 1994. *An Introduction to Systemic Functional Linguistics*. London: Pinter.

Eggins, Suzanne. 2004. *An Introduction to Systemic Functional Linguistics*. London: Continuum.

Elgemark, Anna. 2017. To the Very End. A Contrastive Study of N-Rhemes in English and Swedish Translations. PhD dissertation, University of Gothenburg.

Enkvist, Nils. 1973. Theme dynamics and style: An experiment. *Studia Anglica Posnaniensia* 5(12): 127–135.

Enkvist, Nils. 1978. Coherence, pseudo-coherence, and non-coherence. In *Cohesion and Semantics*, Jan-Ola Östman (ed.), 109–128. Åbo: Åbo Akademi Forskningsinstitut.

Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford: OUP.

Espunya, Anna. 2007. Informativeness and explicit linking in the translation of the English V-ing free adjuncts into Catalan. *Languages in Contrast* 7(2): 143–166.
https://doi.org/10.1075/lic.7.2.04esp

Fabricius-Hansen, Cathrine. 2005. Elusive connectives. A case study on the explicitness dimension of discourse coherence. *Linguistics* 43(1): 17–48. https://doi.org/10.1515/ling.2005.43.1.17

Fabricius-Hansen, Cathrine, Ramm, Wiebke, Solfjeld, Kåre & Behrens, Bergljot. 2005. Coordination, discourse relations, and information packaging - cross-linguistic differences. In *Proceedings of the Symposium on the Exploration and Modelling of Meaning (SEM05)*, Michel Aurnague, Myriam Bras, Anne Le Draoulec & Laure Vieu (eds), 85–93. Biarritz, France.

Fawcett, Robin. 2000. *A Theory of Syntax for Systemic Functional Linguistics* [Current Issues in Linguistic Theory 206]. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.206

Fawcett, Robin. 2007. The many types of "Theme" in English: Their semantic systems and their functional syntax. 1–105. <http://www.cardiff.ac.uk/chri/researchpapers/humanities/papers1-10/4Fawcett.pdf> (4 September 2020).

Field, Andy, Miles, Jeremy & Field, Zoë. 2012. *Discovering Statistics Using R*. London: Sage.

Field, Yvette & Yip, Lee. 1992. A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English. *RELC Journal* 23(1): 15–28. https://doi.org/10.1177/003368829202300102

Filipović, Rudolf. 1974. The use of a corpus in contrastive studies. In *Trends in Kontrastiver Linguistik*, Horst Raabe (ed.), 51–66. Tübingen: Gunther Narr.

Firbas, Jan. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511597817

Firbas, Jan. 1995. A contribution on a panel discussion on rheme. In *Thematic Development of English Texts*, Mohsen Ghadessy (ed.), 213–222. London: Pinter.

Fløttum, Kjersti, Dahl, Trine & Kinn, Torodd. 2006. *Academic Voices: Across Languages and Disciplines* [Pragmatics & Beyond New Series 148]. Amsterdam: John Benjamins.
https://doi.org/10.1075/pbns.148

Flowerdew, John & Forest, Richard. 2015. *Signalling Nouns in Academic English*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139135405

Foolen, Ad. 1991. Polyfunctionality and the semantics of adversative conjunctions. *Multilingua* 10(1): 72–92.

Fraser, Bruce. 1998. Contrastive discourse markers in English. In *Discourse Markers. Description and Theory* [Pragmatics & Beyond New Series 57], Andreas H. Jucker & Yael Ziv (eds), 301–326. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.57.15fra

Fraser, Bruce. 1999. What are discourse markers? *Journal of Pragmatics* 31(7): 931–952. https://doi.org/10.1016/S0378-2166(98)00101-5

Fraser, Bruce & Malamud-Makowski, Monica. 1996. English and Spanish contrastive discourse markers. *Language Sciences* 18(3): 863–881. https://doi.org/10.1016/S0388-0001(96)00052-

Fretheim, Thorstein & Johansson, Stig. 2002. The semantics and pragmatics of the Norwegian concessive marker "likevel": Evidence from the English-Norwegian Parallel Corpus. *Language and Computers* 40(1): 81–101.

Fries, Peter. 1994. On theme, rheme and discourse goals. In *Advances in Written Text Analysis*, Malcolm Coulthard (ed.), 229–249. London: Routledge.

Fries, Peter. 1995. A personal view of theme. In *Thematic Development in English Texts*, Mohsen Ghadessy (ed.), 1–19. London: Pinter.

Gallagher, John. 1995. L'effacement des connecteurs adversatifs et concessifs en français moderne. In *Relations discursives et traduction*, Michel Ballard (ed.), 201–220. Villeneuve d'Ascq: Presses Universitaires de Lille.

Gast, Volker. 2012. Contrastive analysis: Theories and methods. In *Dictionaries of Linguistics and Communication Science: Linguistic Theory and Methodology*, Bernd Kortmann & Johannes Kabatek (eds), Berlin: Mouton de Gruyter.

Gellerstam, Martin. 1996. Translations as a source for cross-linguistic studies. In *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies, Lund 4–5 March 1994*, Karin Aijmer, Bengt Altenberg & Mats Johansson (eds), 53–62. Lund: Lund University Press.

Gilquin, Gaëtanelle. 2010. *Corpus, Cognition and Causative Constructions* [Studies in Corpus Linguistics 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.39

Glaud, Ludivine, Loiseau, Yves & Merlet, Elise. 2015. *Grammaire essentielle du français. Niveau B1*. Paris: Didier.

Gliemann, Marie-Françoise, Bonenfant, Joëlle, Bazelle-Shahmaei, Bernadette & Akyüz, Anne. 2015. *Focus: Grammaire du français*. Paris: Hachette.

Goatly, Andrew. 2004. Corpus linguistics, Systemic Functional Grammar and literary meaning: A critical analysis of Harry Potter and the Philosopher's Stone. *Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies* 46: 115–154. https://doi.org/10.5007/%x

Gómez González, María de los Ángeles. 2001. *The Theme–Topic Interface: Evidence from English* [Pragmatics & Beyond New Series 71] Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.71

González Rodríguez, María José. 2007. On the interpretation of ideology through comment articles: Two views in opinion discourse. *RAEL: Revista Electrónica de Lingüística Aplicada* 1: 49–68.

Granger, Sylviane. In press. Phraséologie et lexicographie bilingue: Apports croisés des corpus monolingues et parallèles. In *Autour de l'énonciation, de la lexicologie, de la morphophonologie et de la contrastivité: Langues, discours, textes et corpus*, Sylvie Hanote & Raluca Nita (eds). Rennes: Presses Universitaires de Rennes.

Granger, Sylviane. 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies, Lund 4–5 March 1994*, Karin Aijmer, Bengt Altenberg & Mats Johansson (eds), 37–51. Lund: Lund University Press.

Granger, Sylviane. 1997. On identifying the syntactic and discourse features of participle clauses in academic English: Native and non-native writers compared. In *Studies in English Language and Teaching*, Jan Aarts, Inge de Mönnink & Herman Wekker (eds), 185–198. Amsterdam: Rodopi.

Granger, Sylviane. 1998. The computer learner corpus: A versatile new source of data for SLA research. In *Learner English on Computer*, Sylviane Granger (ed.), 3–18. London: Longman.

Granger, Sylviane. 2003. The corpus approach: A common way forward for Contrastive Linguistics and Translation Studies. In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, Sylviane Granger, Jacques Lerot & Stephanie Petch-Tyson (eds), 17–29. Amsterdam: Rodopi.

Granger, Sylviane. 2010. Comparable and translation corpora in cross-linguistic research. Design, analysis and applications. *Journal of Shanghai Jiaotong University* 2: 14–21.

Granger, Sylviane. 2014. A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast* 14(1): 58–72. https://doi.org/10.1075/lic.14.1.04gra

Granger, Sylviane. 2018. Tracking the third code: A cross-linguistic corpus-driven approach to metadiscursive markers. In *The Corpus Linguistics Discourse* [Studies in Corpus Linguistics 87], A. Cermakova & Michaela Mahlberg (eds), 185–204. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.87.08gra

Granger, Sylviane & Lefer, Marie-Aude. 2013. Enriching the phraseological coverage of bilingual dictionaries: The respective contribution of monolingual and bilingual corpus data. In *Advances in Corpus-based Contrastive Linguistics. Studies in Honour of Stig Johansson* [Studies in Corpus Linguistics 54], Karin Aijmer & Bengt Altenberg (eds), 157–176. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.54.10gra

Granger, Sylviane & Lefer, Marie-Aude (eds). 2020. *The Complementary Contribution of Comparable and Parallel Corpora to Crosslinguistic Studies. Special issue of Languages in Contrast* 20(2).

Granger, Sylviane, Lerot, Jacques & Petch-Tyson, Stephanie. 2003. Preface. In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, Sylviane Granger, Jacques Lerot & Stephanie Petch-Tyson (eds), 9–13. Amsterdam: Rodopi.

Granger, Sylviane & Tyson, Stephanie. 1996. Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes* 15(1): 17–27. https://doi.org/10.1111/j.1467-971X.1996.tb00089.x

Gray, Bethany. 2015. *Linguistic Variation in Research Articles* [Studies in Corpus Linguistics 71]. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.71

Greenbaum, Sidney. 1969. *Studies in English Adverbial Usage*. London: Longman.

Greenbaum, Sidney. 1988. Syntactic devices for compression in English. In *Essays on the English Language and Applied Linguistics on the Occasion of Gerhard Nickel's 60th Birthday*, Josef Klegraf, Dietrich Nehls & Gerhard Nickel (eds), 3–10. Heidelberg: Julius Groos.

Grégoire, Maïa & Thievenaz, Odile. 2017. *Grammaire progressive du français. Niveau intermédiaire*. Paris: Clé International.

Grellet, Françoise. 1991. *Apprendre à traduire: Typologie d'exercices et de traduction*. Nancy: Presses Universitaires de Nancy.

Grevisse, Maurice & Goosse, André. 1995. *Nouvelle grammaire française: Grammaire*. Bruxelles: De Boeck Supérieur.

Grevisse, Maurice & Goosse, André. 2011. *Le bon usage*. Bruxelles: De Boeck - Duculot.

Gries, Stefan T. 2006. Exploring variability within and between corpora: Some methodological considerations. *Corpora* 1(2): 109–151. https://doi.org/10.3366/cor.2006.1.2.109

Gries, Stefan T. 2009. What is corpus linguistics? *Language and Linguistics Compass* 3(5): 1225–1241. https://doi.org/10.1111/j.1749-818X.2009.00149.x

Gries, Stefan T. 2013. *Statistics for Linguistics with R. A Practical Introduction*. Berlin: Mouton De Gruyter. https://doi.org/10.1515/9783110307474

Grieve, James. 1996. *Dictionary of Contemporary French Connectors*. London: Routledge.

Grote, Brigitte, Lenke, Nils & Stede, Manfred. 1997. Ma(r)king concessions in English and German. *Discourse Processes* 24(1): 87–117. https://doi.org/10.1080/01638539709545008

Guillemin-Flescher, Jacqueline. 1981. *Syntaxe comparée du français et de l'anglais: Problèmes de traduction*. Paris: Ophrys.

Gutwinski, Waldemar. 1976. *Cohesion in Literary Texts. A Study of some Grammatical and Lexical Features of English Discourse*. Berlin: De Gruyter Mouton.
https://doi.org/10.1515/9783111352176

Hadermann, Pascale & Degand, Liesbeth. 2009. Structure narrative et connecteurs temporels en français langue seconde. In *La langue en contexte*, Eva Havu, Juhani Härmä, Mervi Helkkula, Meri Larjavaara & Ulla Tuomarla (eds), 19–34. Helsinki: Société Néophilologique.

Hajičová, Eva. 1994. Topic/Focus and related research. In *The Prague School of Structural and Functional Linguistics* [Linguistic and Literary Studies in Eastern Europe 41], Philip A. Luelsdorff (ed.), 245–275. Amsterdam: John Benjamins.
https://doi.org/10.1075/llsee.41.10haj

Halliday, Michael A. K. 1967. Notes on transitivity and theme in English, Part 2. *Journal of Linguistics* 3(2): 199–244. https://doi.org/10.1017/S0022226700016613

Halliday, Michael A. K. 1970. Language structure and language function. In *New Horizons in Linguistics*, John Lyons (ed.), 140–165. Harmondsworth: Penguin.

Halliday, Michael A. K. 1971. Language in a social perspective. *Educational Review* 23(3): 165–188. https://doi.org/10.1080/0013191710230302

Halliday, Michael A. K. 1977. Text as semantic choice in social contexts. In *Grammars and Descriptions: Studies in Text Theory and Text Analysis*, Teun van Dijk & János Petöfi (eds), 176–225. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110839609-009

Halliday, Michael A. K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Halliday, Michael A. K. 1985. *An Introduction to Functional Grammar*. London: Hodder Arnold.

Halliday, Michael A. K. 1991. Corpus studies and probabilistic grammar. In *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, Karin Aijmer & Bengt Altenberg (eds), 30–43. London: Longman.

Halliday, Michael A. K. 1992. Language as system and language as instance: The corpus as a theoretical construct. In *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4–8 August 1991*, Jan Svartvik (ed.), 61–77. Berlin: Walter de Gruyter.
https://doi.org/10.1515/9783110867275.61

Halliday, Michael A. K. 2005. *Computational and Quantitative Studies*, Jonathan J. Webster (ed.). London: Continuum.

Halliday, Michael A. K. 2006. Afterwords. In *System and Corpus: Exploring Connections*, Geoff Thompson & Susan Hunston (eds), 293–299. London: Equinox.

Halliday, Michael A. K. & Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.

Halliday, Michael A. K. & Hasan, Ruqaiya. 1989. *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective*. Oxford: OUP.

Halliday, Michael A. K. & James, Zoe. 1993. A quantitative study of polarity and primary tense in the English finite clause. In *Techniques of Description: Spoken and Written Discourse*, John Sinclair, Michael Hoey & Gwyneth Fox (eds), 93–128. London: Routledge.

Halliday, Michael A. K. & Matthiessen, Christian M. I. M. 2004. *An Introduction to Functional Grammar*, 3rd edn. London: Routledge.

Halverson, Sandra. 2004. Connectives as a translation problem. In *An International Encyclopedia of Translation Studies*, Werner Koller, Frank Harald, Paul Armin, Norbert Greiner, Jose Lambert & Fritz Paul (eds), 562–572. Berlin: Walter de Gruyter.

Hannay, Mike & Gómez-González, María de los Ángeles. 2012. Thematic parentheticals in Dutch and English. *Linguistics and the Human Sciences* 6: 99–127. https://doi.org/10.1558/lhs.v6i1-3.99

Hansen-Schirra, Silvia & Neumann, Stella. 2003. The challenges of working with multilingual corpora. In *Proceedings of the Workshop on Multilingual Corpora, Linguistics Requirements and Technical Perspectives. Corpus Linguistics Conference 2003*, Stella Neumann & Silvia Hansen-Schirra (eds), 27–34. Lancaster.

Hansen-Schirra, Silvia, Neumann, Stella & Steiner, Erich. 2012. *Cross-linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110260328

Hartnett, Carolyn. 1995. The pit after the theme. In *Thematic Development of English Texts*, Mohsen Ghadessy (ed.), 198–212. London: Pinter.

Hasan, Ruqaiya. 1984. Coherence and cohesive harmony. In *Understanding Reading Comprehension: Cognition, Language, and the Structure of Prose*, James Flood (ed.), 181–219. Newark DE: International Reading Association.

Haselow, Alexander. 2011. Discourse marker and modal particle: The functions of utterance-final then in spoken English. *Journal of Pragmatics* 43(14): 3603–3623. https://doi.org/10.1016/j.pragma.2011.09.002

Haselow, Alexander. 2012. Subjectivity, intersubjectivity and the negotiation of common ground in spoken discourse: Final particles in English. *Language and Communication* 32(3): 182–204. https://doi.org/10.1016/j.langcom.2012.04.008

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3): 663–687. https://doi.org/10.1353/lan.2010.0021

Hasselgård, Hilde. 2004a. Thematic choice in English and Norwegian. *Functions of Language* 11(2): 187–212. https://doi.org/10.1075/fol.11.2.03has

Hasselgård, Hilde. 2004b. The role of multiple themes in cohesion. In *Discourse Patterns in Spoken and Written Corpora* [Pragmatics & Beyond New Series 120], Karin Aijmer & Anna-Brita Stenström (eds), 65–87. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.120.06has

Hasselgård, Hilde. 2010a. Contrastive analysis/contrastive linguistics. In *The Routledge Linguistics Encyclopaedia*, Kirsten Malmkjaer (ed.). London: Routledge.

Hasselgård, Hilde. 2010b. *Adjunct Adverbials in English*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511676253

Hasselgård, Hilde. 2014a. Additive conjunction across languages: "dessuten" and its correspondences in English and French. *Oslo Studies in Language* 6(1): 69–89. https://doi.org/10.5617/osla.691

Hasselgård, Hilde. 2014b. Discourse-structuring functions of initial adverbials in English and Norwegian news and fiction. *Languages in Contrast* 14(1): 73–92. https://doi.org/10.1075/lic.14.1.05has

Hawes, Thomas & Thomas, Sarah. 2012. Theme choice in EAP and media language. *Journal of English for Academic Purposes* 11(3): 175–183. https://doi.org/10.1016/j.jeap.2012.04.005

Hawkins, Roger & Towell, Richard. 2001. *French Grammar and Usage*. London: Routledge.

Hempel, Susanne & Degand, Liesbeth. 2008. Sequencers in different text genres: Academic writing, journalese and fiction. *Journal of Pragmatics* 40(4): 676–693. https://doi.org/10.1016/j.pragma.2007.02.001

He, Qingshun & Wen, Binli. 2017. A corpus-based study of textual metaphor in English. *Australian Journal of Linguistics* 37(3): 265–285. https://doi.org/10.1080/07268602.2017.1298393

He, Qingshun & Yang, Bingjun. 2015. *Absolute Clauses in English from the Systemic Functional Perspective: A Corpus-Based Study*. Berlin: Springer. https://doi.org/10.1007/978-3-662-46367-3

Herbst, Thomas, Schmid, Hans-Jörg & Faulhaber, Susen. 2014. From collocations and patterns to constructions – An introduction. In *Constructions, Collocations, Patterns*, Thomas Herbst, Hans-Jörg Schmid & Susen Faulhaber (eds), 1–8. Berlin: De Gruyter Mouton.

Herriman, Jennifer. 2011. N-rhemes in English problem–solution texts. *English Text Construction* 4(1): 29–53. https://doi.org/10.1075/etc.4.1.03her

Herriman, Jennifer & Bostrom Aronsson, Mia. 2009. Themes in Swedish advanced learners' writing in English. In *Corpora and Language Teaching* [Studies in Corpus Linguistics, 33], Karin Aijmer (ed.), 101–120. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.33.11her

Hervey, Sándor & Higgins, Ian. 1992. *Thinking Translation: A Course in Translation Method, French-English*. London: Routledge.

Hervey, Sándor & Higgins, Ian. 2002. *Thinking French Translation: A Course in Translation Method: French to English*. London: Routledge. https://doi.org/10.4324/9780203417973

Hoarau, Lucie. 1997. *Étude contrastive de la coordination en français et en anglais*. Paris: Ophrys.

Hoek, Jet, Evers-Vermeul, Jacqueline & Sanders, Ted J. M. 2018. Segmenting discourse: Incorporating interpretation into segmentation? *Corpus Linguistics and Linguistic Theory* 14(2): 357–386. https://doi.org/10.1515/cllt-2016-0042

Hoek, Jet & Zufferey, Sandrine. 2015. Factors influencing the implicitation of discourse relations across languages. In *Conference Proceedings 11th Joint ACL - ISO Workshop on Interoperable Semantic Annotation, London*, Harry Bunt (ed.), 39–45. Tilburg: Tilburg Centre for Cognition and Communication.

Hoek, Jet, Zufferey, Sandrine, Evers-Vermeul, Jacqueline & Sanders, Ted J. M. 2017. Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics* 121(Supplement C): 113–131. https://doi.org/10.1016/j.pragma.2017.10.010

Hoey, Michael. 1991. *Patterns of Lexis in Text*. Oxford: OUP.

Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Holtz, Monica. 2007. Corpus-based analysis of verb/noun collocations in interdisciplinary registers. In *Proceedings of the Corpus Linguistics Conference 2007. University of Birmingham, UK 27–30 July 2007*, Matthew Davies, Paul Rayson, Susan Hunston & Danielsson Pernilla (eds). <https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2007/14Paper.pdf> (4 September 2020).

Huddleston, Rodney & Pullum, Geoffrey. 2002. *The Cambdrige Grammar of the English Language*. Cambridge: CUP. https://doi.org/10.1017/9781316423530

Hundt, Marianne & Mair, Christian. 1999. "Agile" and "uptight" genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2): 221–242. https://doi.org/10.1075/ijcl.4.2.02hun

Hunston, Susan. 2013. Systemic functional linguistics, corpus linguistics, and the ideology of science. *Text & Talk* 33(4-5): 614–640. https://doi.org/10.1515/text-2013-0028

Hunston, Susan & Francis, Gill. 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English* [Studies in Corpus Linguistics 4]. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.4

Hyland, Ken. 1998. Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics* 30(4): 437–455. https://doi.org/10.1016/S0378-2166(98)00009-5

Hyland, Ken & Jiang, Feng (Kevin). 2017. Is academic writing becoming more informal? *English for Specific Purposes* 45: 40–51. https://doi.org/10.1016/j.esp.2016.09.001

Iruskieta, Mikel, da Cunha, Iria & Taboada, Maite. 2014. A qualitative comparison method for rhetorical structures: identifying different discourse structures in multilingual corpora. *Language Resources and Evaluation* 49(2): 263–309. https://doi.org/10.1007/s10579-014-9271-6

Ivir, Vladimir. 1983. A translation-based model of contrastive analysis. *Jyväskylä Cross-Language Studies* 9: 171–178.

Izutsu, Mitsuko Narita. 2008. Contrast, concessive, and corrective: Toward a comprehensive study of opposition relations. *Journal of Pragmatics* 40(4): 646–675.
https://doi.org/10.1016/j.pragma.2007.07.001

Jacobson, Sven. 1964. *Adverbial Positions in English*. Uppsala: Studentbok.

James, Carl. 1980. *Contrastive Analysis*. London: Longman.

Janicki, Karol. 1990. On the tenability of the notion "pragmatic equivalence" in contrastive analysis. In *Further Insights into Contrastive Analysis* [Linguistic and Literary Studies in Eastern Europe 30], Jacek Fisiak (ed.), 47–54. Amsterdam: John Benjamins.

Jaszczolt, Katarzyna. 2003. On translating "what is said": Tertium comparationis in contrastive semantics and pragmatics. In *Meaning Through Language Contrast*, Vol. 2, Katarzyna Jaszczolt & Ken Turner (eds), 441–462. Amsterdam: John Benjamins.
https://doi.org/10.1075/pbns.100.26jas

Johansson, Stig. 1998. On the role of corpora in cross-linguistic research. In *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, Stig Johansson & Signe Oksefjell (eds), 3–24. Amsterdam: Rodopi.

Johansson, Stig. 2003. Contrastive linguistics and corpora. In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, Sylviane Granger, Jacques Lerot & Stephanie Petch-Tyson (eds), 31–44. Amsterdam: Rodopi.

Johansson, Stig. 2007. *Seeing Through Multilingual Corpora: On the Use of Corpora in Contrastive Studies* [Studies in Corpus Linguistics 26]. Amsterdam: John Benjamins.
https://doi.org/10.1075/scl.26

Johansson, Stig. 2012. Cross-linguistic perspectives. In *English Corpus Linguistics: Crossing Paths*, Merjä Kytö (ed.), 45–68. Amsterdam: Rodopi.

Johansson, Stig & Hofland, Knud. 1994. Towards an English-Norwegian parallel corpus. In *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora, Zürich 1993*, Udo Fries, Gunnel Tottie & Peter Schneider (eds), 25–37. Amsterdam: Rodopi.

Johansson, Stig & Lysvåg, Per. 1986. *Understanding English Grammar: An Overview*. Oslo: Universitetsforlaget.

Kaplan, Robert B. 1966. Cultural thought patterns in inter-cultural education. *Language Learning* 16(1-2): 1–20. https://doi.org/10.1111/j.1467-1770.1966.tb00804.x

Károly, Krisztina. 2017. *Aspects of Cohesion and Coherence in Translation: The Case of Hungarian-English News Translation*. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.134

Knott, Alistair. 1996. A Data-driven Methodology for Motivating a Set of Coherence Relations. PhD dissertation, University of Edinburgh.

König, Ekkehard. 2012. Contrastive linguistics and language comparison. *Languages in Contrast* 12(1): 3–26. https://doi.org/10.1075/lic.12.1.02kon

König, Ekkehard & Siemund, Peter. 2000. Causal and concessive clauses: formal and semantic relations. In *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives*,

Elizabeth Couper-Kuhlen & Bernd Kortmann (eds), 341–360. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110219043.4.341

Kortmann, Bernd. 1991. *Free Adjuncts and Absolutes in English: Problems of Control and Interpretation*. London: Routledge.

Kruger, Haidee. 2017. The effects of editorial intervention: Implications for studies of the features of translated language. In *Empirical Translation Studies: New Methodological and Theoretical Traditions*, Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (eds), 113–155. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110459586-005

Kruger, Haidee & van Rooy, Bertus. 2012. Register and the features of translated language. *Across Languages and Cultures* 13(1): 33–65. https://doi.org/10.1556/Acr.13.2012.1.3

Krzeszowski, Tomasz P. 1984. Tertium comparationis. In *Contrastive Linguistics: Prospects and Problems*, Jacek Fisiak (ed.), 301–312. Berlin: Walter de Gruyter.

Krzeszowski, Tomasz P. 1990. *Contrasting Languages: The Scope of Contrastive Linguistics*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110860146

Kunz, Kerstin, Degaetano-Ortlieb, Stefania, Lapshinova-Koltunski, Ekaterina, Menzel, Katrien & Steiner, Erich. 2017. English-German contrasts in cohesion and implications for translation. In *Empirical Translation Studies: New Methodological and Theoretical Traditions*, Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (eds), 265–311. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110459586-010

Kunz, Kerstin & Lapshinova-Koltunski, Ekaterina. 2014. Cohesive conjunctions in English and German: Systemic contrasts and textual differences. In *Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora*, Lieven Vandelanotte, Kristin Davidse, Caroline Gentens & Ditte Kimps (eds), 229–262. Amsterdam: Rodopi. https://doi.org/10.1163/9789401211130_012

Kunz, Kerstin & Lapshinova-Koltunski, Ekaterina. 2015. Cross-linguistic analysis of discourse variation across registers. *Nordic Journal of English Studies* 14(1): 258–288. https://doi.org/10.35360/njes.347

Kunz, Kerstin & Steiner, Erich. 2012. Towards a comparison of cohesive reference in English and German: System and text. In *Contrastive Discourse Analysis: Functional and Corpus Perspectives*, Maite Taboada, Susana Doval Suarez & Elsa Gonzalez Alvarez (eds), 208–239. Sheffield: Equinox.

Kunz, Kerstin & Steiner, Erich. 2013. Cohesive substitution in English and German: A contrastive and corpus-based perspective. In *Advances in Corpus-based Contrastive Linguistics: Studies in Honour of Stig Johansson* [Studies in Corpus Linguistics 54] Karin Aijmer & Bengt Altenberg (eds), 201–232. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.54.12kunz

Kurokawa, David, Goutte, Cyril & Isabelle, Pierre. 2009. Automatic detection of translated text and its impact on Machine Translation. *Proceedings of the Twelfth Machine Translation Summit, Ottawa, August 26–30*, 81–88. <http://www.cs.cmu.edu/~dkurokaw/publications/MTS-2009-Kurokawa.pdf> (4 September 2020).

Lado, Robert. 1957. *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor MI: University of Michigan Press.

Lakoff, Robin. 1971. If's, and's and but's about conjunction. In *Studies in Linguistic Semantics*, Charles J. Fillmore & D. Terence Langndoen (eds), 3–114. New York NY: Holt.

Lambrecht, Knud. 1988. Presentational cleft constructions in spoken French. In *Clause Combining in Grammar and Discourse* [Typological Studies in Language 18], John Haiman & Sandra A. Thompson (eds), 135–179. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.18.08lam

Lambrecht, Knud. 2010. Constraints on subject-focus mapping in French and English. In *Comparative and Contrastive Studies of Information Structure* [Linguistik Aktuell/Linguistics Today 165], Carsten Breul & Edward Göbbel (eds), 77–100. Amsterdam: John Benjamins. https://doi.org/10.1075/la.165.04lam

Lamiroy, Béatrice & Van Belle, William. 1995. Connectives of contrast and concession in Dutch and French. *Leuvense Bijdragen: Tijdschrift voor Germaanse Filologie* 84(3): 397–418.

Lamiroy, Béatrice & Vanderbauwhede, Gudrun. 2016. Connecteurs et linguistique contrastive. Les marqueurs de discours "en effet", "en fait", "de fait", "en réalité" et leurs pendants néerlandais. In *Connexion et indexation. Ces liens qui lient le texte*, Laure Sarda, Denis Vigier & Bernard Combettes (eds), 195–209. Lyon: ENS Editions.

Lapshinova-Koltunski, Ekaterina & Kunz, Kerstin. 2014. Annotating cohesion for multilingual analysis. *Proceedings of the 10th Joint ACL–ISO Workshop on Interoperable Semantic Annotation*, 57–64. Reykjavik: ACL/ISO.

Lapshinova-Koltunski, Ekaterina, Nedoluzhko, Anna & Kunz, Kerstin. 2015. Across languages and genres: Creating a universal annotation scheme for textual relations. *Proceedings of The 9th Linguistic Annotation Workshop*, 168–177. https://doi.org/10.3115/v1/W15-1620

Larsson, Tove. 2017a. The importance of, it is important that or importantly? The use of morphologically related stance markers in learner and expert writing. *International Journal of Corpus Linguistics* 22(1): 57–84. https://doi.org/10.1075/ijcl.22.1.03lar

Larsson, Tove. 2017b. A functional classification of the introductory it pattern: Investigating academic writing by non-native-speaker and native-speaker students. *English for Specific Purposes* 48: 57–70. https://doi.org/10.1016/j.esp.2017.06.001

Lauridsen, Karen. 1996. Text corpora and contrastive linguistics: which type of corpus for which type of analysis? In *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies, Lund 4–5 March 1994*, Karin Aijmer, Bengt Altenberg & Mats Johansson (eds), 63–71. Lund: Lund University Press.

Lavid, Julia. 2010. Contrasting choices in clause-initial position in English and Spanish: A corpus-based analysis. In *Thresholds and Potentialities of Systemic Functional Linguistics: Multilingual, Multimodal and Other Specialised Discourses*, Elizabeth Swain (ed.), 49–68. Trieste: EUT Edizioni Università di Trieste.

Lavid, Julia, Arús, Jorge & Zamorano-Mansilla, Juan Rafael. 2011. *Systemic Functional Grammar of Spanish*. London: Continuum.

Laviosa, Sara. 2009. Universals. In *Routledge Encyclopedia of Translation Studies*, Mona Baker & Gabriela Saldanha (eds), 306–310. London: Routledge.

Le Draoulec, Anne & Bras, Myriam. 2006. Quelques candidats au statut de 'connecteur temporel'. *Cahiers de Grammaire* 30: 219–237.

Leech, Geoffrey & Svartvik, Jan. 2002. *A Communicative Grammar of English*. London: Pearson Education.

Lefer, Marie-Aude. 2009. Exploring Lexical Morphology across Languages: A Corpus-based Study of Prefixation in English and French Writing. PhD dissertation, Université Catholique de Louvain.

Lefer, Marie-Aude & Cartoni, Bruno. 2011. Prefixes in contrast: Towards a meaning-based contrastive methodology for lexical morphology. *Languages in Contrast* 11(1): 87–105. https://doi.org/10.1075/lic.11.1.07lef

Lefer, Marie-Aude & Grabar, Natalia. 2015. Super-creative and over-bureaucratic: A cross-genre corpus-based study on the use and translation of evaluative prefixation in TED talks and EU parliamentary debates. *Across Languages and Cultures* 16: 187–208. https://doi.org/10.1556/084.2015.16.2.3

Lefer, Marie-Aude & Vogeleer, Svetlana (eds). 2014. *Genre- and Register-related Discourse Features in Contrast. Special issue of Languages in Contrast* 14(1).

Lefeuvre, Florence. 1999. *La phrase averbale en français*. Paris: L'Harmattan.

Le Goffic, Pierre. 1994. *Grammaire de la phrase française*. Paris: Hachette.

Lehmann, Christian. 1988. Towards a typology of clause linkage. In *Clause Combining in Grammar and Discourse* [Typological Studies in Language 18], John Haiman & Sandra A. Thompson (eds), 181–226. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.18.09leh

Lenker, Ursula. 2010. *Argument and Rhetoric. Adverbial Connectors in the History of English*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110216066

Lenker, Ursula. 2011. A focus on adverbial connectors: Connecting, partitioning and focusing attention in the history of English. *Varieng: Studies in Variation, Contacts and Change in English* 8. <http://www.helsinki.fi/varieng/series/volumes/08/lenker/> (4 September 2020).

Lenker, Ursula. 2014. Knitting and splitting information: Medial placement of linking adverbials in the history of English. In *Contact, Variation and Change in the History of English* [Studies in Language Companion Series 159], Simone E. Pfenninger, Olga Timofeeva, Anne-Christine Gardner, Alpo Honkapohja, Marianne Hundt & Daniel Schreier (eds), 11–38. Amsterdam: John Benjamins.

Leroux, Agnès. 2012. La relation inter-énonciative et le marquage syntaxique des relations de cause: Étude contrastive anglais-français. *CORELA* HS-10. https://doi.org/10.4000/corela.2429

Levshina, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.195

Lewis, Diana. 2005. Mapping adversative coherence relations in English and French. *Languages in Contrast* 5(1): 33–48. https://doi.org/10.1075/lic.5.1.05lew

Lewis, Diana. 2006. Contrastive analysis of adversative relational markers using comparable corpora. In *Pragmatic Markers in Contrast*, Karin Aijmer & Anne-Marie Simon-Vandenbergen (eds), 139–153. Oxford: Elsevier.

Lewis, Diana. 2009. Markers of concession in a contrastive perspective: evidence from an English/French comparable corpus. In *Corpora and Discourse - and Stuff: Papers in Honour of Karin Aijmer*, Rhonwen Bowen, Mats Mobärg & Sölve Ohlander (eds), 189–198. Göteborg: Acta Universitatis Gothoburgensis.

Liu, Dilin. 2008. Linking adverbials: An across-register corpus study and its implications. *International Journal of Corpus Linguistics* 13(4): 491–518. https://doi.org/10.1075/ijcl.13.4.05liu

Loock, Rudy. 2018. Using non-standard word order you should! A corpus-based approach to avoiding standardized word order in translated French. In *Using Corpora in Contrastive and Translation Studies Conference*, 5th edn [CECL Papers 1], Sylviane Granger, Marie-Aude Lefer & Laura Aguiar de Souza Penha Marion (eds), 115–116. Louvain-la-Neuve: Université catholique de Louvain.

Lores, Rosa. 2004. On RA abstracts: From rhetorical structure to thematic organisation. *English for Specific Purposes* 23(3): 280–302. https://doi.org/10.1016/j.esp.2003.06.001

Love, Alison. 2004. Drawing on (a lot of) "Given": One aspect of theme choice in newspaper editorials. In *Text and Texture. Systemic Functional Viewpoints on the Nature and Structure of Text*, David Banks (ed.). Paris: L'Harmattan.

Lynch, Jack. 2007. *The English Language: A User's Guide*. Newburyport MA: Focus.

Macken, Lieve, De Clercq, Orphée & Paulussen, Hans. 2011. Dutch Parallel Corpus: A balanced copyright-cleared parallel corpus. *Meta* 56(2): 374–390. https://doi.org/10.7202/1006182ar

Mair, Christian & Hundt, Marianne. 1995. Why is the progressive becoming more frequent in English? A corpus-based investigation of language change in progress. *Zeitschrift für Anglistik und Amerikanistik* 43(2): 111–122.

Mann, William & Thompson, Sandra. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8: 243–281. https://doi.org/10.1515/text.1.1988.8.3.243

Mann, William & Thompson, Sandra. 1992. Relational discourse structure: A comparison of approaches to structuring text by "contrast." In *Language in Context: Essays for Robert E. Longacre*, Shin Ja Hwang & William Merrifield (eds), 19–45. Dallas TX: Summer Institute of Linguistics.

Martel, Guylaine. 1993. Les connecteurs contre-argumentatifs en anglais, en français et en espagnol: Une question d'usage. *Langues et Linguistique* 19: 151–165.

Martin, James R. 1983. Conjunction: The logic of English text. In *Micro and Macro Connexity of Texts*, János Petöfi & Emel Sözer (eds), 1–72. Hamburg: Helmut Buske.

Martin, James R. 1992. *English Text: System and Structure*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.59

Martin, James R. 2001. Cohesion and texture. In *The Handbook of Discourse Analysis*, Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton (eds), 35–53. London: Blackwell.

Martin, James R. & Rose, David. 2007. *Working with Discourse: Meaning beyond the Clause*. London: Continuum.

Mason, Ian. 1998. Discourse connectives, ellipsis and markedness. In *The Pragmatics of Translation*, Leo Hickey (ed.), 170–186. Clevedon: Multilingual Matters.

Mason, Ian. 2001. Translator behaviour and language usage: Some constraints on contrastive studies. *Hermes* 26: 65–80.

Matthiessen, Christian M. I. M. 1992. Interpreting the textual metafunction. In *Advances in Systemic Linguistics: Recent Theory and Practice*, Martin Davies & Louise Ravelli (eds), 37–81. London: Pinter.

Matthiessen, Christian M. I. M. 2002. Combining clauses into clause complexes: A multi-faceted view. In *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson*, Joan Bybee & Michael Noonan (eds), 235–319. Amsterdam: John Benjamins. https://doi.org/10.1075/z.110.13mat

Matthiessen, Christian M. I. M. 2006. Frequency profiles of some basic grammatical systems: An interim report. In *System and Corpus: Exploring Connections*, Geoff Thompson & Susan Hunston (eds), 103–142. London: Equinox.

Matthiessen, Christian M. I. M. & Thompson, Sandra. 1988. The structure of discourse and "subordination." In *Clause Combining in Grammar and Discourse* [Typological Studies in Language 18], John Haiman & Sandra Thompson (eds), 275–329. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.18.12mat

Mauranen, Anna. 1999. Will "translationese" ruin a contrastive study? *Languages in Contrast* 2(2): 161–185. https://doi.org/10.1075/lic.2.2.03mau

Mauranen, Anna. 2008. Universal tendencies in translation. In *Incorporating Corpora: The Linguist and the Translator*, Gunilla M. Anderman & Margaret Rogers (eds), 32–48. Clevedon: Multilingual Matters.

McDonald, Daniel & Woodward-Kron, Robyn. 2016. Member roles and identities in online support groups: Perspectives from corpus and systemic functional linguistics. *Discourse & Communication* 10(2): 157–175. https://doi.org/10.1177/1750481315615985

McEnery, Tony & Wilson, Andrew. 2001. *Corpus Linguistics: An Introduction*. Edinburgh: EUP.

McEnery, Tony & Xiao, Richard. 2007. Parallel and comparable corpora: What are they up to? In *Incorporating Corpora: Translation and the Linguist*, Gunilla M. Anderman & Margaret Rogers (eds), 17–21. Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853599873-005

McEnery, Tony, Xiao, Richard & Tono, Yukio. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.

Meyer, David, Zeileis, Achim, Hornik, Kurt, Gerber, Florian & Friendly, Michael. 2017. *vcd: Visualizing Categorical Data*. <https://CRAN.R-project.org/package=vcd> (4 September 2020).

Meyer, Thomas. 2011. Disambiguating temporal-contrastive connectives for machine translation. In *Proceedings of the ACL-HLT 2011 Student Session*, 46–51. Portland OR.

Meyer, Thomas, Popescu-Belis, Andrei, Zufferey, Sandrine & Cartoni, Bruno. 2011. Multilingual annotation and disambiguation of discourse connectives for Machine Translation. In *Proceedings of the SIGDIAL 2011 Conference (SIGDIAL '11)*, 194–203. Stroudsburg PA: Association for Computational Linguistics.

Mikhailov, Mikhail & Cooper, Robert. 2016. *Corpus Linguistics for Translation and Contrastive Studies: A Guide for Research*. London: Routledge. https://doi.org/10.4324/9781315624570

Miller, Donna & Johnson, Jane. 2013. "Register-idiosyncratic" evaluative choice in Congressional debate. In *Systemic Functional Linguistics: Exploring Choice*, Lise Fontaine, Tom Bartlett & Gerard O'Grady (eds), 417–431. Cambridge: CUP. https://doi.org/10.1017/CBO9781139583077.026

Miller, Donna R. & Johnson, Jane. 2014. Evaluative phraseological choice and speaker party/gender: A corpus-assisted comparative study of "register-idiosyncratic" meaning in Congressional debate. In *Evaluation in Context* [Pragmatics & Beyond New Series 242], Geoff Thompson & Laura Alba-Juez (eds), 345–366. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.242.17mil

Miltsakaki, Eleni, Dinesh, Nikhil, Prasad, Rashmi, Joshi, Aravind & Webber, Bonnie. 2005. Experiments on sense annotations and sense disambiguation of discourse connectives. In Proceedings of the 4th Workshop on Treebanks and Linguistic Theories (TLT), 9–10 December 2005, Barcelona, Spain.

Miltsakaki, Eleni, Prasad, Rashmi, Joshi, Aravind & Bonnie, Webber. 2004. The Penn Discourse Treebank. In Proceedings of the 4th Language Resources and Evaluation Conference. Lisbon, Portugal.

Morel, Mary-Annick. 1996. *La concession en français*. Paris: Ophrys.

Morel, Mary-Annick & Danon-Boileau, Laurent. 1998. *Grammaire de l'intonation. L'exemple du français oral*. Paris: Ophrys.

Moreno, Ana I. 1998. The explicit signaling of premise-conclusion sequences in research articles: A contrastive framework. *Text* 18(4): 545–585. https://doi.org/10.1515/text.1.1998.18.4.545

Mortier, Liesbeth & Degand, Liesbeth. 2009. Adversative discourse markers in contrast: The need for a combined corpus approach. *International Journal of Corpus Linguistics* 14(3): 3–301. https://doi.org/10.1075/ijcl.14.3.03mor

Muller, Philippe, Vergez-Couret, Marianne, Prévot, Laurent, Asher, Nicholas, Farah, Benamara, Bras, Myriam, Le Draoulec, Anne & Vieu, Laure. 2012. *Manuel d'annotation en relations de discours du projet ANNODIS*. Rapport technique. Toulouse: CLLE-ERSS.

Murray, John D. 1997. Connectives and narrative text: The role of continuity. *Memory & Cognition* 25(2): 227–236. https://doi.org/10.3758/BF03201114

Mwinlaaru, Isaac N. & Xuan, Winfred Wenhui. 2016. A survey of studies in systemic functional language description and typology. *Functional Linguistics* 3. https://doi.org/10.1186/s40554-016-0030-4

Nesbitt, Christopher & Plum, Guenther. 1988. Probabilities in a systemic-functional grammar: The clause complex in English. In *New Developments in Systemic Linguistics: Theory and Application*, Vol. 2, Robin Fawcett & David Young (eds), 6–38. London: Pinter.

Neumann, Stella. 2010. Quantitative register analysis across languages. In *Thresholds and Potentialities of Systemic Functional Linguistics: Multilingual, Multimodal and Other Specialised Discourses*, Elizabeth Swain (ed.), 85–113. Trieste: EUT Edizioni.

Neumann, Stella. 2012. Register-induced properties of translations. In *Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German*, Silvia Hansen-Schirra, Stella Neumann & Erich Steiner (eds), 191–209. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110260328.191

Neumann, Stella. 2013. *Contrastive Register Variation. A Quantitative Approach to the Comparison of English and German*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110238594

Neumann, Stella. 2014. Cross-linguistic register studies: Theoretical and methodological considerations. *Languages in Contrast* 14(1): 35–57. https://doi.org/10.1075/lic.14.1.03neu

Newmark, Peter. 1988. *A Textbook of Translation*. New York NY: Prentice Hall.

Nome, Astrid & Hobæk Haff, Marianne. 2011. Une analyse contrastive de "donc." *Oslo Studies in Language* 3(1): 47–67. https://doi.org/10.5617/osla.184

O'Donnell, Michael. 2009. The UAM CorpusTool: Software for corpus annotation and exploration. In *Applied Linguistics Now: Understanding Language and Mind / La lingüística aplicada actual: Comprendiendo el lenguaje y la mente*, Carmen M. Bretones Callejas, José Francisco Fernández Sánchez, José Ramón Ibáñez Ibáñez, María Elena García Sánchez, Mª Enriqueta Cortés de los Ríos, Sagrario Salaberri Ramiro, Mª Soledad Cruz Martínez, Nobel Perdú Honeyman, Blasina Cantizano Márquez (eds), 1433–1448. Almeria: Universidad de Almería.

Osborne, John. 2008. Adverb placement in post-intermediate learner English: A contrastive study of learner corpora. In *Linking up Contrastive and Learner Corpus Research*, Gaëtanelle Gilquin, Szilvia Papp & María Belén Díez-Bedmar (eds), 127–146. Amsterdam: Rodopi. https://doi.org/10.1163/9789401206204_006

Øverås, Linn. 1998. In search of the third code: An investigation of norms in literary translation. *Meta* 43(4): 557–570. https://doi.org/10.7202/003775ar

Oversteegen, Leonoor. 1997. On the pragmatic nature of causal and contrastive connectives. *Discourse Processes* 24(1): 51–85. https://doi.org/10.1080/01638539709545007

Pander Maat, Henk & Degand, Liesbeth. 2001. Scaling causal relations and connectives in terms of speaker involvement. *Cognitive Linguistics* 12(3): 211–245. https://doi.org/10.1515/cogl.2002.002

Paquot, Magali. 2010. *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. London: Continuum.

Pekelder, Jan. 2010. Le tertium comparationis en linguistique contrastive. Problèmes et méthodes. *Linguistica Pragensia* 20(1): 22–37. https://doi.org/10.2478/v10017-010-0002-6

Petukhova, Volha & Bunt, Harry. 2009. Towards a multidimensional semantics of discourse markers in spoken dialogue. In *Proceedings of the Eighth International Conference on Computational Semantics (IWCS-8 '09)*, 157–168. Stroudsburg PA: Association for Computational Linguistics.

Pit, Mirna. 2007. Cross-linguistic analyses of backward causal connectives in Dutch, German and French. *Languages in Contrast* 7(1): 53–82. https://doi.org/10.1075/lic.7.1.04pit

Plonsky, Luke & Oswald, Frederick. 2014. How big is "big"? Interpreting effect sizes in L2 research. *Language Learning* 64(4): 878–912. https://doi.org/10.1111/lang.12079

Plum, Guenther & Cowling, Ann. 1987. Social constraints on grammatical variables: Tense choice in English. In *Language Topics: Essays in Honour of Michael Halliday*, Vol. 2, Ross Steele & Terry Threadgold (eds), 281–292. Amsterdam: John Benjamins. https://doi.org/10.1075/z.lt2.66plu

Poncharal, Bruno. 2005. Etude contrastive de la structuration du discours en anglais et en français dans des textes de sciences humaines. In *Actes du Colloque "D'une langue à l'autre": Besançon, 5-6-7 septembre 2002*. Daniel Lebaud (ed.), 287–302. Besançon: Presses Universitaires de Franche-Comté.

Poncharal, Bruno. 2007. Cohérence discursive en anglais et en français: Fonction des connecteurs dans la traduction. In *Les connecteurs, jalons du discours*, Agnès Celle, Stéphane Gresset & Ruth Huart (eds), 117–136. Bern: Peter Lang.

Prasad, Rashmi, Dinesh, Nikhil, Lee, Alan, Miltsakaki, Eleni, Robaldo, Livio, Joshi, Aravind & Webber, Bonnie. 2008. The Penn Discourse Treebank 2.0. In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC). Marrakech, Morocco.

Prasad, Rashmi, Joshi, Aravind & Webber, Bonnie. 2010. Realization of discourse relations by other means: Alternative lexicalizations. In *Coling2010: Poster Volume*, 1023–1031. Beijing.

Prince, Ellen. 1997. On the functions of left-dislocation in English discourse. In *Directions in Functional Linguistics* [Studies in Language Companion Series 36], Akio Kamio (ed.), 117–143. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.36.08pri

Quillard, Geneviève. 1997. Étude de certaines différences dans l'organisation collective des textes pragmatiques anglais et français. *Babel* 43(4): 313–330. https://doi.org/10.1075/babel.43.4.04qui

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. 1972. *A Grammar of Contemporary English*. London: Longman.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Ramón García, Noelia. 2002. Contrastive Linguistics and Translation Studies interconnected: The corpus-based approach. In *Linguistica Antverpiensia New Series – Themes in Translation Studies* 1, Leona van Vaerenbergh (ed.), 393–406. Antwerp: University of Antwerp.

Ravelli, Louise. 1995. A dynamic perspective: implications for metafunctional interaction and an understanding of Theme. In *On Subject and Theme: A Discourse Functional Perspective* [Current Issues in Linguistic Theory 40], Ruqaiya Hasan & Peter Fries (eds), 187–234. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.118.07rav

R Development Core Team. 2008. *R: A Language and Eenvironment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org> (28 September 2020).

Régent, Odile. 1980. Approche comparative des discours de spécialité pour l'entraînement à l'anglais écrit. *Mélanges Pédagogiques* 1: 117–135.

Régent, Odile. 1992. Pratiques de communication en médecine: Contextes anglais et français. *Langages* 26(105): 66–75. https://doi.org/10.3406/lgge.1992.1624

Régent, Odile. 1994. L'article scientifique: Un produit culturel. *ASp. la revue du GERAS* 5–6: 55–59. https://doi.org/10.4000/asp.4017

Rey, Joëlle. 1999. Approche argumentative des textes scientifiques: La traduction de "or" en espagnol. *Meta: Journal des Traducteurs* 44(3): 411–428. https://doi.org/10.7202/003946ar

Riegel, Martin, Pellat, Jean-Christophe & Rioul, René. 2001. *Grammaire méthodique du français*. Paris: PUF.

Rivelin-Constantin, Eve. 1992. La thématisation en français et en anglais: Une étude contrastive. In *Linguistique contrastive et traduction*, Vol. 1, Jacqueline Guillemin-Flescher (ed.), 159–204. Gap: Ophrys.

Rørvik, Sylvi & Egan, Thomas. 2013. Connectors in the argumentative writing of Norwegian novice writers. In *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead. Proceedings of the First Learner Corpus Research Conference (LCR 2011)*, Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (eds), 401–410. Louvain-la-Neuve: Presses Universitaires de Louvain.

Rossette, Fiona. 2007. Connecteurs, enchaînements discursifs et lisibilité en anglais: Quelques pistes de réflexion. In *Les connecteurs, jalons du discours*, Agnès Celle, Stéphane Gresset & Ruth Huart (eds), 11–42. Bern: Peter Lang.

Rossette, Fiona. 2009. Thème, conjonction et cohésion: Corrélations entre les différents composants de la métafonction textuelle en français. In *La linguistique systémique fonctionnelle et la langue française*, David Banks, Simon Eason & Janet Ormrod (eds), 9–42. Paris: L'Harmattan.

Roze, Charlotte, Danlos, Laurence & Muller, Philippe. 2012. LEXCONN: A French Lexicon of Discourse Connectives. *Discours* 10. https://doi.org/10.4000/discours.8645

Rubattel, Christian. 1982. De la syntaxe des connecteurs pragmatiques. *Cahiers de Linguistique Française* 4: 37–61.

Rudolph, Elisabeth. 1996. *Contrast: Adversative and Concessive Relations and Their Expressions in English, German, Spanish, Portuguese on Sentence and Text Level*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110815856

Ruquet, Michel, Quoy-Bodin, Jean-Luc & Cayol, Micheline. 1991. *Comment dire? Raisonner à la française*. Paris: Clé International.

Rysová, Magdalena & Rysová, Katerina. 2014. The centre and periphery of discourse connectives. In *28th Pacific Asia Conference on Language, Information and Computation*, 452–459. Phuket, Thailand.

Sajavaara, Kari. 1996. New challenges for contrastive linguistics. In *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies, Lund 4–5 March 1994*, Karin Aijmer, Bengt Altenberg & Mats Johansson (eds), 17–36. Lund: Lund University Press.

Salkie, Raphael. 2008. How can lexicographers use a translation corpus? In *Proceedings of the International Symposium on Using Corpora in Contrastive and Translation Studies (UCCTS 2008)*, Richard Xiao, Lianzhen He & Ming Yue (eds). Zhejiang University, Hangzhou.

Salkie, Raphael & Oates, Sarah Louise. 1999. Contrast and concession in French and English. *Languages in Contrast* 2(1): 27–56. https://doi.org/10.1075/lic.2.1.04sal

Salkoff, Morris. 1999. *A French-English Grammar: A Contrastive Grammar on Translational Principles*. Amsterdam: John Benjamins. https://doi.org/10.1075/lis.22

Sanders, Ted J. M. 1997. Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes* 24(1): 119–147. https://doi.org/10.1080/01638539709545009

Sanders, Ted J. M. 2005. Coherence, causality and cognitive complexity in discourse. In *Proceedings/Actes SEM-05, First International Symposium on the Exploration and Modelling of Meaning*, 105–114. Biarritz, France.

Sanders, Ted J.M, Demberg, Vera, Hoek, Jet, Scholman, Merel, Asr, Fatemeh, Zufferey, Sandrine & Evers-Vermeul, Jacqueline. 2018. Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory* (published

online ahead of print 22 May). <https://www.degruyter.com/view/j/cllt.ahead-of-print /cllt-2016-0078/cllt-2016-0078.xml> (4 September 2020). https://doi.org/10.1515/cllt-2016-0078

Sanders, Ted J.M & Spooren, Wilbert. 2007. Discourse and text structure. In *The Oxford Handbook of Cognitive Linguistics*, Dirk Geeraerts & Hubert Cuyckens (eds), 916–943. Oxford: OUP.

Sanders, Ted J.M, Spooren, Wilbert & Noordman, Leo. 1992. Toward a taxonomy of coherence relations. *Discourse Processes* 15: 1–35. https://doi.org/10.1080/01638539209544800

Schiffrin, Deborah. 2006. Discourse. In *An Introduction to Language and Linguistics*, Ralph W. Fasold & Jeff Connor-Linton (eds), 169–203. Cambridge: CUP.

Schiffrin, Deborah, Tannen, Deborah & Hamilton, Heidi. 2015. Introduction. In *The Handbook of Discourse Analysis*, Deborah Tannen, Heidi Hamilton & Deborah Schiffrin (eds), 1–7. Oxford: Wiley-Blackwell.

Schilperoord, Joost & Verhagen, Arie. 1998. Conceptual dependency and the clausal structure of discourse. In *Discourse and Cognition. Bridging the Gap*, Jean-Pierre Koenig (ed.), 141–163. Stanford CA: CSLI.

Schmied, Josef. 2009. Contrastive corpus studies. In *Corpus Linguistics. An International Handbook*, Vol. 2, Anke Lüdeling & Merja Kytö (eds), 1140–1159. Berlin: Mouton de Gruyter.

Schourup, Lawrence. 1999. Discourse markers. *Lingua* 107(3): 227–265. https://doi.org/10.1016/S0024-3841(96)90026-1

Scott, Mike. 2012. *WordSmith Tools 6*. Liverpool: Lexical Analysis Software.

Sharoff, Serge. 2017. Corpus and systemic functional linguistics. In *The Routledge Handbook of Systemic Functional Linguistics*, Tom Bartlett & Gerard O'Grady (eds), 533–546. London: Routledge.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.

Sinclair, John. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Smith, Raoul & Frawley, William. 1983. Conjunctive cohesion in four English genres. *Text* 3(4): 347–374. https://doi.org/10.1515/text.1.1983.3.4.347

Spooren, Wilbert & Degand, Liesbeth. 2010. Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory* 6(2): 241–266. https://doi.org/10.1515/cllt.2010.009

Staples, Shelley, Egbert, Jesse, Biber, Douglas & Conrad, Susan. 2015. Register variation. A corpus approach. In *The Handbook of Discourse Analysis*, Deborah Tannen, Heidi Hamilton & Deborah Schiffrin (eds), 505–525. Oxford: Wiley-Blackwell.

Steen, Gerard. 2003. Conversationalization in discourse: Stylistic changes in editorials of The Times between 1950 and 2000. In *Determination of Information and Tenor in Texts: Multidisciplinary Approaches to Discourse*, Luuk Lagerwerf, Wilbert Spooren & Liesbeth Degand (eds), 115–124. Amsterdam & Munster: Stichting Neerlandistiek & Nodus Publikationen.

Stoye, Hélène. 2014. *Les connecteurs contenant des prépositions en français: Profils sémantiques et pragmatiques en synchronie et diachronie*. Berlin: Walter de Gruyter.

Struck, Herman R. 1965. The myth about initial conjunctions. *The English Journal* 54(1): 42–44. https://doi.org/10.2307/810944

Sweetser, Eve. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511620904

Taboada, Maite. 2006. Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics* 38(4): 567–592. https://doi.org/10.1016/j.pragma.2005.09.010

Taboada, Maite. 2009. Implicit and explicit coherence relations. In *Discourse, of Course*, Jan Renkema (ed.), 127–140. Amsterdam: John Benjamins. https://doi.org/10.1075/z.148.13tab

Taboada, Maite & Gómez-González, María de los Ángeles. 2012. Discourse markers and coherence relations: Comparison across markers, languages and modalities. *Linguistics and the Human Sciences* 6: 17–41. https://doi.org/10.1558/lhs.v6i1-3.17

Taglicht, Josef. 1984. *Message and Emphasis: On Focus and Scope in English*. London: Longman.

Takagaki, Yumi. 2011. *Les plans d'organisation textuelle en français et en japonais: De la rhétorique contrastive à la linguistique textuelle*. Rouen: Editions universitaires européennes.

Tankó, Gyula. 2004. The use of adverbial connectors in Hungarian university students' argumentative essays. In *How to Use Corpora in Language Teaching* [Studies in Corpus Linguistics 12], John Sinclair (ed.), 157–181. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.12.13tan

Tanskanen, Sanna-Kaisa. 2006. *Collaborating Towards Coherence: Lexical Cohesion in English Discourse* [Pragmatics & Beyond New Series 146] Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.146

Taylor, Charlotte. 2008. What is corpus linguistics? What the data says. *ICAME Journal* 32: 179–200.

Teich, Elke. 2003. *Cross-Linguistic Variation in System and Text*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110896541

Teich, Elke. 2009. Linguistic computing. In *Continuum Companion to Systemic Functional Linguistics*, Michael A. K. Halliday & Jonathan J. Webster (eds), 113–127. London: Continuum.

Teich, Elke. 2013. Choices in analysing choice: Methods and techniques for register analysis. In *Systemic Functional Linguistics: Exploring Choice*, Lise Fontaine, Tom Bartlett & Gerard O'Grady (eds), 417–431. Cambridge: CUP. https://doi.org/10.1017/CBO9781139583077.025

Teich, Elke, Eckart, Richard & Holtz, Monica. 2006. Systemic Functional corpus resources: Issues in development and deployment. In *Proceedings of the Fifth International Treebanks and Linguistic Theories Conference (TLT 06)*, 247–258. Prague.

Teubert, Wolfgang. 1996. Comparable or parallel corpora? *International Journal of Lexicography* 9(3): 238–264. https://doi.org/10.1093/ijl/9.3.238

Thompson, Geoff. 2014. *Introducing Functional Grammar*. London: Routledge. https://doi.org/10.4324/9780203785270

Thompson, Geoff & Hunston, Susan. 2006a. System and corpus: two traditions with a common ground. In *System and Corpus: Exploring Connections*, Geoff Thompson & Susan Hunston (eds), 1–14. London: Equinox.

Thompson, Geoff & Hunston, Susan. 2006b. *System and Corpus: Exploring Connections*. London: Equinox.

Thompson, Geoff & Hunston, Susan. 2008. Theme, subject and the unfolding of text. In *Text Type and Texture*, Gail Forey & Geoff Thompson (eds), 45–69. London: Equinox.

Thompson, Geoff & Zhou, Jianglin. 2001. Evaluation and organization in text: The structuring role of evaluative disjuncts. In *Evaluation in Text: Authorial Stance and the Construction of Discourse*, Susan Hunston & Geoff Thompson (eds), 121–141. Oxford: OUP.

Tierney, Robert & Mosenthal, James. 1983. Cohesion and textual coherence. *Research in the Teaching of English* 17(3): 215–229.

Tizón-Couto, David. 2012. *Left Dislocation in English: A Functional-Discoursal Approach*. Bern: Peter Lang.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work* [Studies in Corpus Linguistics 6]. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.6

Touratier, Christian. 2006. Que faut-il entendre par "connecteur"? In *La connexion et les connecteurs. La phrase existentielle*, Christian Touratier & Jean-Marie Merle (eds), 19–40. Université de Provence. Provence.

Travis, Catherine & Torres Cacoullos, Rena. 2012. Discourse syntax. In *The Handbook of Hispanic Linguistics*, José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds), 653–672. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118228098.ch30

Trévise, Anne. 1986. Is it transferable, topicalization? In *Crosslinguistic Influence in Second Language Acquisition*, Eric Kellerman & Michael Sharwood Smith (eds), 86–206. New York NY: Pergamon Institute of English.

Tucker, Gordon. 2006. Systemic incorporation: On the relationship between corpus and systemic functional grammar. In *System and Corpus: Exploring Connections*, Geoff Thompson & Susan Hunston (eds), 81–102. London: Equinox.

Tucker, Gordon. 2009. Towards a lexicogrammar of war: A corpus-based systemic functional investigation of the French lexical item "guerre." In *La linguistique systémique fonctionnelle et la langue française*, David Banks, Simon Eason & Janet Ormrod (eds), 195–222. Paris: L'Harmattan.

Vandepitte, Sonia & De Sutter, Gert. 2013. Contrastive linguistics and translation studies. In *Handbook of Translation Studies*, Vol. 4, Yves Gambier & Luc van Doorslaer (eds), 36–41. Amsterdam: John Benjamins. https://doi.org/10.1075/hts.4.con4

Van der Auwera, Johan. 2012. From contrastive linguistics to linguistic typology. *Languages in Contrast* 12(1): 69–86. https://doi.org/10.1075/lic.12.1.05auw

Van Dijk, Teun. 1979. Pragmatic connectives. *Journal of Pragmatics* 3: 447–456. https://doi.org/10.1016/0378-2166(79)90019-5

Van Hoof, Henri. 1989. *Traduire l'anglais: Théorie et pratique*. Louvain-la-Neuve: Duculot.

Van de Voorde, Katrien. 1992. De deux à trois "mais": Essai de vérification des approches d'Anscombre et Ducrot et de Blumenthal. *Travaux de Linguistique* 24: 57–81.

Verhagen, Arie. 2000. Concession implies causality, though in some other space. In *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives*, Elizabeth Couper-Kuhlen & Bernd Kortmann (eds), 361–380. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110219043.4.361

Verhagen, Arie. 2001. Subordination and discourse segmentation revisited, or: Why matrix clauses may be more dependent than complements. In *Text Representation: Linguistic and Psychological Aspects* [Human Cognitive Processing 8], Ted J. M. Sanders, Joost Schilperoord & Wilbert Spooren (eds), 337–357. Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.8.18ver

Vinay, Jean-Paul & Darbelnet, Jean. 1995. *Comparative Stylistics of French and English: A Methodology for Translation*. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.11

Virtanen, Tuija. 1992. *Discourse functions of Adverbial Placement in English: Clause-initial Adverbials of Time and Place in Narratives and Procedural Place Descriptions*. Åbo: Åbo Akademis Förlag.

Virtanen, Tuija. 2005. "Polls and surveys show": Public opinion as a persuasive device in editorial discourse. In *Persuasion across Genres: A Linguistic Approach* [Pragmatics & Beyond New Series 130], Elena Halmari & Tuija Virtanen (eds), 153–180. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.130.10vir

Wang, Jianxin. 2011. Contrastive Connectors in English and Chinese: A Corpus-Based Study. PhD dissertation, University of Auckland.

Westin, Ingrid. 2002. *Language Change in English Newspaper Editorials*. Amsterdam: Rodopi. https://doi.org/10.1163/9789004334007

White, Lydia. 1991. Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Interlanguage Studies Bulletin* 7(2): 133–161. https://doi.org/10.1177/026765839100700205

Widdowson, Henry G. 1978. *Teaching Language as Communication*. Oxford: OUP.

Willems, Dominique, Defrancq, Bart, Colleman, Timothy & Noël, Dirk. 2004. *Contrastive Analysis in Language. Identifying Linguistic Units of Comparison*. Houndmills: Palgrave Macmillan.

Wu, Canzhong. 2009. Corpus-based research. In *Continuum Companion to Systemic Functional Linguistics*, Michael A. K. Halliday & Jonathan J. Webster (eds), 128–142. London: Continuum.

Xiao, Richard. 2010a. How different is translated Chinese from native Chinese? A corpus-based study of translation universals. *International Journal of Corpus Linguistics* 15(1): 5–35. https://doi.org/10.1075/ijcl.15.1.01xia

Xiao, Richard. 2010b. *Using Corpora in Contrastive and Translation Studies*. Newcastle upon Tyne: Cambridge Scholars.

Xiao, Richard & Yue, Ming. 2009. Using corpora in translation studies: The state of the art. In *Contemporary Corpus Linguistics*, Paul Baker (ed.), 237–262. London: Continuum.

Xi, Yan. 2010. Cohesion studies in the past 30 years: Development, application and chaos. *Language, Society and Culture* 31: 139–147.

Zimmerman, Donald W. 2003. A warning about the large-sample Wilcoxon-Mann-Whitney test. *Understanding Statistics* 2(4): 267–280. https://doi.org/10.1207/S15328031US0204_03

Zinn, Jens & McDonald, Daniel. 2018. *Risk in The New York Times (1987–2014). A Corpus-based Exploration of Sociological Theories*. London: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-64158-4

Zinsser, William K. 2001. *On Writing Well, 25th Anniversary: The Classic Guide to Writing Nonfiction*. New York NY: Collins.

Ziv, Yael. 1994. Left and right dislocations: Discourse functions and anaphora. *Journal of Pragmatics* 22(6): 629–645. https://doi.org/10.1016/0378-2166(94)90033-7

Zufferey, Sandrine. 2012. "Car, parce que, puisque" revisited: Three empirical studies on French causal connectives. *Journal of Pragmatics* 44(2): 138–153. https://doi.org/10.1016/j.pragma.2011.09.018

Zufferey, Sandrine. 2016. Discourse connectives across languages: Factors influencing their explicit or implicit translation. *Languages in Contrast* 16(2): 264–279. https://doi.org/10.1075/lic.16.2.05zuf

Zufferey, Sandrine & Cartoni, Bruno. 2012. English and French causal connectives in contrast. *Languages in Contrast* 12(2): 232–250. https://doi.org/10.1075/lic.12.2.06zuf

Zufferey, Sandrine & Degand, Liesbeth. 2017. Annotating the meaning of discourse connectives in multilingual corpora. *Corpus Linguistics and Linguistic Theory* 13(2): 399–422. https://doi.org/10.1515/cllt-2013-0022

Zufferey, Sandrine, Degand, Liesbeth, Popescu-Belis, Andrei & Sanders, Ted J. M. 2012. Empirical validations of multilingual annotation schemes for discourse relations. In *Proceedings of the 8th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, Harry Bunt (ed.), 77–84. Pisa, Italy.

Zufferey, Sandrine & Gygax, Pascal. 2016. The role of perspective shifts for processing and translating discourse relations. *Discourse Processes* 53(7): 532–555.
https://doi.org/10.1080/0163853X.2015.1062839

Zufferey, Sandrine, Mak, Willem, Degand, Liesbeth & Sanders, Ted J. M. 2015. Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research* 31(3): 389–411. https://doi.org/10.1177/0267658315573349

Appendices

Appendix 1: List of newspapers and academic journals included in the corpus

Table 1. List of newspapers included in Mult-Ed

LANGUAGE	NEWSPAPER	TOTAL NUMBER OF TEXTS	TOTAL NUMBER OF WORDS
ENGLISH	*The Economist*	63	58,853
	The Financial Times	369	232,310
	The Guardian	1,357	653,221
	The Independent	622	292,110
	The Observer	243	145,550
	The Telegraph	1,396	552,150
	The Times	130	72,546
FRENCH	*Le Figaro*	1,170	576,837
	Libération	1,818	693,144
	Le Monde	1,138	601,235
	Le Nouvel Observateur	277	224,959

Table 2. List of academic journals included in LOCRA

LANGUAGE	DISCIPLINE	JOURNAL	TOTAL NUMBER OF TEXTS	TOTAL NUMBER OF WORDS
ENGLISH	ANTHROPOLOGY	*American Anthropologist*	13	107,648
		Annual Review of Anthropology	19	113,312
		Cultural Anthropology	17	101,800
		Ethnology	14	90,115
	EDUCATION	*American Educational Research Journal*	8	84,080
		Instructional Science	12	90,197
		International Journal of Educational Development	10	73,604
		Teaching and Teacher Education	10	77,521
		Higher Education	14	83,325

(Continued)

Table 2. (*Continued*)

Language	Discipline	Journal	Total number of texts	Total number of words
	Political science	*Foreign Policy Analysis*	12	86,922
		International Organization	9	73,502
		Journal of Common Market Studies	13	80,187
		Journal of Public Policy	11	81,998
		Public Administration Review	11	79,307
	Psychology	*Behavior Therapy*	13	75,355
		Journal of Applied Psychology	8	81,075
		Journal of Experimental Psychology	9	87,202
		Journal of Personality and Social Psychology	7	70,443
		Psychological Review	6	90,820
	Sociology	*American Journal of Sociology*	9	107,170
		The British Journal of Sociology	16	102,938
		Community, Work and Family	14	88,625
		Current Sociology	16	107,677
French	Anthropology	*L'Homme*	9	80,133
		L'Anthropologie	15	100,995
		Ethnologie française	25	133,438
		Civilisations - Revue internationale d'anthropologie et de sciences humaines	12	86,551
	Education	*Revue française de pédagogie*	13	78,154
		Les sciences de l'éducation - pour l'ère nouvelle	14	80,487
		Revue internationale de pédagogie de l'enseignement supérieur	14	81,889
		Recherche et Formation	19	83,162
		Recherches et éducations	19	79,650
	Political science	*Revue française de science politique*	9	82,844
		Politix: revue de science sociale du politique	10	83,100

Table 2. (*Continued*)

Language	Discipline	Journal	Total number of texts	Total number of words
		Raisons politiques – études de pensée	9	57,888
		Politique Européenne	12	88,504
		Revue internationale de politique comparée	13	92,650
	Psychology	*Année psychologique*	11	81,794
		Revue européenne de psychologie appliquée	10	78,798
		Psychologie française	12	83,794
		Revue internationale de psychologie sociale	12	81,102
		Pratiques psychologiques	13	81,635
	Sociology	*Revue française de sociologie*	9	99,961
		L'année sociologique	13	103,283
		Cahiers internationaux de sociologie	15	91,726
		Sociologies pratiques	25	102,151

Appendix 2: Individual syntactic patterning of English and French conjunctive markers of contrast

1. Clause types

Table 1. Raw frequencies of use of individual English conjunctive adjuncts across clause types in Mult-Ed

Conjunctive adjunct	Frequency in finite clauses	Frequency in non-finite clauses	Frequency in verbless clauses	Total frequency in Mult-Ed
All the same	20	0	2	22
Anyway	81	3	0	84
At the same time	36	3	0	39
By contrast	91	0	0	91
Conversely	14	1	0	15
Even so	83	0	0	83

(*Continued*)

Table 1. (*Continued*)

Conjunctive adjunct	Frequency in finite clauses	Frequency in non-finite clauses	Frequency in verbless clauses	Total frequency in Mult-Ed
However	1,203	1	4	1,208
In contrast	17	0	0	17
Instead	317	52	0	369
Meanwhile	48	0	0	48
Nevertheless	167	0	2	169
Nonetheless/none the less	133	0	0	133
On the contrary	42	0	0	42
On the other	17	2	1	20
On the other hand	78	0	0	78
Rather	56	1	1	58
Still	297	8	0	305
Though	325	0	3	328
Yet	1,533	10	4	1,547
TOTAL	**4,558**	**81**	**17**	**4,656**

Table 2. Raw frequencies of use of individual English subordinators across clause types in Mult-Ed

Subordinator	Frequency in finite clauses	Frequency in non-finite clauses	Frequency in verbless clauses	Total frequency in Mult-Ed
Albeit	1	3	38	42
Although	535	11	44	490
Even if	384	8	8	400
Even though	172	0	0	172
Though	403	23	80	506
Whereas	54	0	0	54
While	948	159	46	1,153
TOTAL	**2,497**	**204**	**216**	**2,917**

Table 3. Raw frequencies of use of individual French conjunctive adjuncts across clause types in Mult-Ed

Conjunctive adjunct	Frequency in finite clauses	Frequency in non-finite clauses	Frequency in verbless clauses	Phrasal-clause uses	Total frequency in Mult-Ed
À l'inverse	15	5	0	0	20
Au contraire	229	20	2	1	252
Bien au contraire	23	0	0	0	23
Cependant	266	7	9	0	282
De l'autre	18	3	26	0	47
De toute façon	25	0	2	0	27
De toute manière	11	0	0	0	11
En même temps	10	1	0	0	11
En revanche	222	3	5	0	230
Malgré tout	35	8	0	0	43
Néanmoins	104	3	2	1	110
Or	355	0	1	0	356
Plutôt	110	5	15	0	130
Pourtant	918	36	19	56	1,029
Quand même	16	0	1	0	17
Tout de même	72	8	8	0	88
Toutefois	121	12	3	0	136
TOTAL	**2,550**	**111**	**93**	**58**	**2,812**

Table 4. Raw frequencies of use of individual French subordinators across clause types in Mult-Ed

Subordinator	Frequency in finite clauses	Frequency in non-finite clauses	Frequency in verbless clauses	Total frequency in Mult-Ed
Alors même que	55	0	0	55
Alors qu*	547	0	1	548
Bien qu*	45	12	5	62
Même s*	618	0	2	620
Quand bien même	19	0	0	19
Quoique	5	6	17	28
Si/s'	656[#]	0	0	656[#]
Tandis qu*	106	0	0	106
Tout en	0	116	0	116
TOTAL	**2,051**	**134**	**25**	**2,210**

2. Rank status

Table 5. Raw frequencies of use of individual English conjunctive adjuncts across ranks in Mult-Ed

Conjunctive adjunct	Frequency in main clauses	Frequency in hypotactic clauses	Frequency in embedded clauses	Frequency in minor clauses	Total frequency in Mult-Ed
All the same	20	0	0	2	22
Anyway	38	30	16	0	84
At the same time	35	4	0	0	39
By contrast	90	1	0	0	91
Conversely	13	0	2	0	15
Even so	83	0	0	0	83
However	1,198	4	1	5	1,208
In contrast	17	0	0	0	17
Instead	302	36	31	0	369
Meanwhile	48	0	0	0	48
Nevertheless	159	3	5	2	169
Nonetheless/none the less	131	1	1	0	133
On the contrary	41	1	0	0	42
On the other	16	1	2	1	20
On the other hand	71	7	0	0	78
Rather	48	0	4	6	58
Still	270	23	12	0	305
Though	325	0	0	3	328
Yet	1,523	7	13	4	1,547
TOTAL	**4,428**	**118**	**87**	**23**	**4,656**

Table 6. Raw frequencies of use of individual English subordinators across ranks in Mult-Ed

Subordinator	Frequency in hypotactic clauses	Frequency in embedded clauses	Frequency in minor clauses	Total frequency in Mult-Ed
Albeit	42	0	0	42
Although	586	0	4	490
Even if	398	0	2	400

Table 6. (*Continued*)

Subordinator	Frequency in hypotactic clauses	Frequency in embedded clauses	Frequency in minor clauses	Total frequency in Mult-Ed
Even though	172	0	0	172
Though	501	1	4	506
Whereas	54	0	0	54
While	1,151	0	2	1,153
TOTAL	**2,904**	**1**	**12**	**2,917**

Table 7. Raw frequencies of use of individual French conjunctive adjuncts across ranks in Mult-Ed

Conjunctive adjunct	Frequency in main clauses	Frequency in hypotactic clauses	Frequency in embedded clauses	Frequency in minor clauses	Phrasal-clausal uses	Total frequency in Mult-Ed
À l'inverse	12	5	2	1	0	20
Au contraire	198	35	11	7	1	252
Bien au contraire	21	2	0	0	0	23
Cependant	261	8	1	12	0	282
De l'autre	19	1	0	27	0	47
De toute façon	15	7	2	3	0	27
De toute manière	8	2	1	0	0	11
En même temps	9	1	1	0	0	11
En revanche	211	7	6	6	0	230
Malgré tout	26	10	7	0	0	43
Néanmoins	96	7	2	4	1	110
Or	355	0	0	1	0	356
Plutôt	97	13	7	13	0	130
Pourtant	807	116	28	22	56	1,029
Quand même	12	4	0	1	0	17
Tout de même	54	21	2	11	0	88
Toutefois	116	17	0	3	0	136
TOTAL	**2,317**	**256**	**70**	**111**	**58**	**2,812**

Table 8. Raw frequencies of use of individual French subordinators across ranks in Mult-Ed

Subordinator	Frequency in hypotactic clauses	Frequency in embedded clauses	Frequency in minor clauses	Total frequency in Mult-Ed
Alors même que	53	0	2	55
Alors qu*	519	0	29	548
Bien qu*	62	0	0	62
Même s*	570	0	50	620
Quand bien même	17	0	2	19
Quoique	27	0	1	28
Si/s'	656#	0	0	656#
Tandis qu*	101	0	5	106
Tout en	113	0	3	116
TOTAL	**2,118**	**0**	**92**	**2,210**

Appendix 3.   Non-pruned Classification and Regression Trees emerging from the analysis of English and French conjunctive adjunct placement
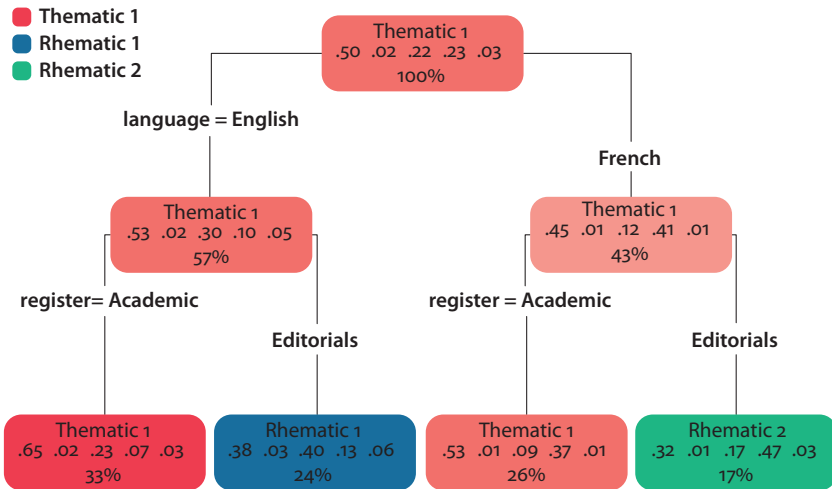
Figure 1. Non-pruned Classification Tree for conjunctive adjunct placement as a function of language and register[1]

1.   Note that in this case, the non-pruned tree is in fact identical to the pruned one. This means that the cross-validation stage did not identify any divisions which were too specific to
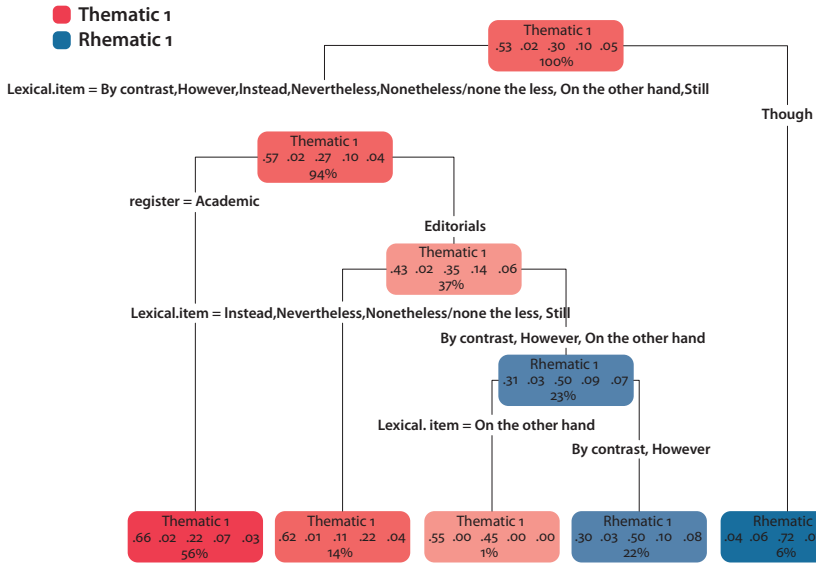
Figure 2. Non-pruned Classification Tree for conjunctive adjunct placement as a function of register and lexis in English
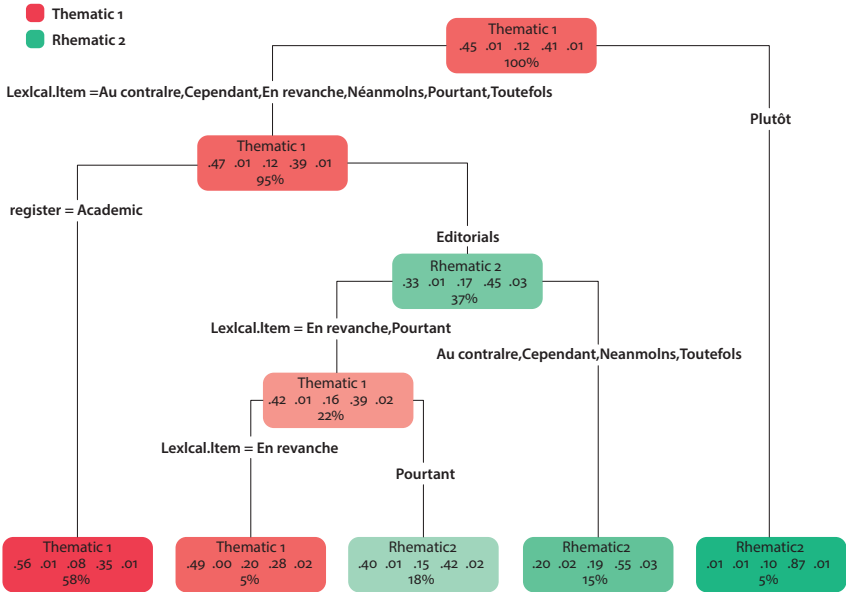
Figure 3. Non-pruned Classification Tree for conjunctive adjunct placement as a function of register and lexis in French

my own data set to also be generalisable to other, similar data sets.

Index

Situated at the interface between corpus linguistics and Systemic Functional Linguistics, this volume focuses on conjunctive markers expressing contrast in English and French. The frequency and placement patterns of the markers are analysed using large corpora of texts from two written registers: newspaper editorials and research articles. The corpus study revisits the long-standing but largely unsubstantiated claim that French requires more explicit markers of cohesive conjunction than English and shows that the opposite is in fact the case. Novel insights into the placement preferences of English and French conjunctive markers are provided by a new approach to theme and rheme that attaches more importance to the rheme than previous studies. The study demonstrates the significant benefits of a combined corpus and Systemic Functional Linguistics approach to the cross-linguistic analysis of cohesion.