

STUDIES IN
LANGUAGE
COMPANION
SERIES 197

Tense, Aspect, Modality, and Evidentiality

Crosslinguistic perspectives

Edited by
Dalila Ayoun
Agnès Celle
Laure Lansari

JOHN BENJAMINS PUBLISHING COMPANY

Copyright © 2015, John Benjamins Publishing Company. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair use permitted under U.S. or applicable copyright law.

Tense, Aspect, Modality, and Evidentiality

Studies in Language Companion Series (SLCS)

ISSN 0165-7763

This series has been established as a companion series to the periodical *Studies in Language*.

For an overview of all books published in this series, please see <http://benjamins.com/catalog/slcs>

Founding Editor

Werner Abraham
University of Vienna / University of Munich

Editors

Werner Abraham
University of Vienna / University of Munich

Elly van Gelderen
Arizona State University

Editorial Board

Bernard Comrie
University of California, Santa Barbara

William Croft
University of New Mexico

Östen Dahl
University of Stockholm

Gerrit J. Dimmendaal
University of Cologne

Ekkehard König
Free University of Berlin

Christian Lehmann
University of Erfurt

Elisabeth Leiss
University of Munich

Marianne Mithun
University of California, Santa Barbara

Heiko Narrog
Tohoku University

Johanna L. Wood
University of Aarhus

Debra Ziegeler
University of Paris III

Volume 197

Tense, Aspect, Modality, and Evidentiality. Crosslinguistic perspectives
Edited by Dalila Ayoun, Agnès Celle and Laure Lansari

Tense, Aspect, Modality, and Evidentiality

Crosslinguistic perspectives

Edited by

Dalila Ayoun

University of Arizona

Agnès Celle

Paris Diderot University & University of Colorado

Laure Lansari

Paris Diderot University

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/slcs.197

Cataloging-in-Publication Data available from Library of Congress.

ISBN 978 90 272 0096 9 (HB)

ISBN 978 90 272 6390 2 (E-BOOK)

© 2018 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

Table of contents

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Preface | VII |
| CHAPTER 1 | |
| Introduction: On the gradience of TAM-E categories <i>Dalila Ayoun, Agnès Celle and Laure Lansari</i> | 1 |
| CHAPTER 2 | |
| A quantitative perspective on modality and future tense in French and German <i>Annalena Hütsch</i> | 19 |
| CHAPTER 3 | |
| The temporal uses of French <i>devoir</i> and Estonian <i>pidama</i> ('must') <i>Anu Treikelder and Marri Amon</i> | 41 |
| CHAPTER 4 | |
| The competition between the present conditional and the prospective imperfect in French over the centuries: First results <i>Jacques Bres, Sascha Diwersy and Giancarlo Luxardo</i> | 65 |
| CHAPTER 5 | |
| Evidentiality and the TAM systems in English and Spanish: A cognitive and cross-linguistic perspective <i>Juana I. Marín Arrese</i> | 83 |
| CHAPTER 6 | |
| Expressing sources of information, knowledge and belief in English and Spanish informative financial texts <i>Marta Carretero and Yolanda Berdasco-Gancedo</i> | 109 |
| CHAPTER 7 | |
| Evidentiality and epistemic modality in Old Catalan: A diachronic cognitive approach to the semantics of modal verbs <i>Andreu Sentí</i> | 145 |

| | | |
|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|-----|
| CHAPTER 8 | | |
| 'I think': An enunciative and corpus-based perspective | | 165 |
| | <i>Graham Ranger</i> | |
| CHAPTER 9 | | |
| Embedding evidence in Tagalog and German: On two types of evidentials | | 185 |
| | <i>Jennifer Tan and Johannes Mursell</i> | |
| CHAPTER 10 | | |
| Questions as indirect speech acts in surprise contexts | | 213 |
| | <i>Agnès Celle</i> | |
| CHAPTER 11 | | |
| Non-finiteness, complementation and evidentiality: The Lithuanian <i>Accusativus cum Participio</i> in a cross-linguistic perspective | | 239 |
| | <i>Aurelija Usonienė and Nigel Vincent</i> | |
| CHAPTER 12 | | |
| The perfect in Avar and Andi: Cross-linguistic variation among two closely-related East Caucasian languages | | 261 |
| | <i>Samira Verhees</i> | |
| CHAPTER 13 | | |
| The different grammars of event singularisation: A cross-linguistic corpus study | | 281 |
| | <i>Eric Corre</i> | |
| CHAPTER 14 | | |
| Phraseological usage patterns of past tenses: A corpus-driven look on French <i>passé composé</i> and <i>imparfait</i> | | 309 |
| | <i>Oliver Wicher</i> | |
| CHAPTER 15 | | |
| Path scales: Directed-motion verbs, prepositions and telicity in European Portuguese | | 335 |
| | <i>António Leal, Fátima Oliveira and Purificação Silvano</i> | |
| Name index | | 357 |
| Subject index | | 363 |

Preface

This volume presents a selection of the papers that were originally presented at the *Tense, Aspect, Modality, Evidentiality: Comparative, Cognitive, Theoretical and Applied Perspectives* conference that took place on November 17–18, 2016 at *Université Paris Diderot* guided by a desire to bring together scholars from all over the world who are contributing to an already rich body of research. And indeed we received numerous submissions from scholars in 28 different countries, targeting 34 different languages, including less commonly studied languages such as Swahili, Xhosa or Turkish.

The languages represented in this volume are mainly Indo-European languages, more specifically Romance languages (French, Spanish, Old Catalan, European Portuguese), Germanic languages (English, German), a Baltic language (Lithuanian), a Slavic language (Russian). However, several contributions are also devoted to the comparative analysis of lesser-studied languages such as Finno-Ugric languages (Hungarian and Estonian), the Avar-Andic branch of the East Caucasian language family (Avar and Andi), and an Austronesian language (Tagalog). These contributions stress temporal, aspectual and modal variation either within a language family (Avar and Andi, Spanish and English) or from a cross-language perspective (Estonian and French, Hungarian, Russian, English and French, Tagalog and German).

While couched within different theoretical frameworks, the contributions are all firmly empirically-grounded, based either on elicited data or on corpus data. Some chapters rely on large diachronic corpora and provide new qualitative insight on the evolution of TAM systems through quantitative methods, while others carry out a collostructional analysis of past-tensed verbs using inferential statistics to explore the lexical grammar of verbs. The common goal of all the chapters is to uncover semantic regularities and variation in the TAM systems of the languages under study by taking a close look at context. Such a fine-grained approach contributes to our understanding of the TAM systems from a typological perspective.

We are deeply grateful to the following colleagues who served as anonymous reviewers: Juana Marín Arrese, Gilles Authier, Sophie Azzopardi, Johan Van der Auwera, Cécile Barbet, Fernando Bermúdez, Philippe Bourdin, Jacques Bres, Laurel Brinton, Mike Child, Llorenc Comajoan, Denis Creissels, Berit Gehrke, Montserrat Gonzalez, Laurent Gosselin, Axel Holvoet, Paul Larreya, Brenda Laca, Diana

Lewis, Josep Martines, Helle Metslang, Gunther Kaltenböck, Timur Maisak, Eric McCready, Jean-Marie Merle, Laura Michaelis, Tanja Mortelmans, Dirk Noël, Rea Peltoa, Maïa Ponsonnet, Graham Ranger, Henriette De Swart, Ye Tian, Anastasios Tsangalidis, Carl Vettters, Marion Weerning, Björn Wiemer, Juan Zamorano.

We also acknowledge the research laboratories and institutions for their support in the organization of the conference: CLILLAC-ARP EA 3967 (Paris Diderot), LATTICE UMR 8094 (Ecole Normale Supérieure, CNRS, Université Paris 3 Sorbonne Nouvelle), Labex Empirical Foundations of Linguistics, Labex TransferS, Région Ile de France, Université Sorbonne Paris Cité (Invitation Longue Durée).

Introduction

On the gradience of TAM-E categories

Dalila Ayoun, Agnès Celle and Laure Lansari

University of Arizona / Université Paris Diderot and University of Colorado /
Université Paris Diderot

1. Categorization in tense, aspect, modality and evidentiality: Theoretical issues and state of the art

The volume is divided into three parts that reflect crucial connections between the categories of tense, aspect, modality and evidentiality (TAM-E) and epitomize the difficulty in drawing clear lines between those categories.

1.1 Futurity, modality and conditionals

Part 1 deals with future time reference and anteriority expressed by the future tense, modal verbs and the conditional. How to represent time has long been a subject of inquiry in philosophy as well as in linguistics. While acknowledging the necessity of an observer for time to emerge, the Aristotelian tradition represents time as a continuous line going from left to right, that is, from the past to the future. This linear representation of time was first associated with an objective conception of time that was consistent with the time measurement of physical processes such as the rotation of the earth and the oscillation of a pendulum. As pointed out by Fleischman (1982) as well as Manning, Cassel & Cassel (2013) among others, the contribution of St. Augustine on time and memory marked a significant change by introducing the concept of subjective time, thereby shifting time representation to the center of human consciousness and cognition. Along this line of thinking, past and future only exist if they are present in our consciousness. Past time is reconstructed through memory and future time is anticipated from the experiential present. It is important to note that anticipation is committed to memory.

Linguistically, the present of things to come corresponds to a future embedded in the present (Fleischman 1982). The linguistic tradition has been strongly influenced by the Aristotelian conception of time. Poutsma (1922: 7), Jespersen (1924: 257), Comrie (1985: 2) advocate the view that time is linear. However, formal studies going back to Prior (1967) have acknowledged the branching structure of the future. More recently, neurolinguistic approaches (van Wassenhove 2017) have stressed that objective time cannot be mapped to subjective time, since our perception of time relies on our brain dynamics. In any case, both formal and neurolinguistic approaches concur that the future can only be envisaged from the present.

Turning to tense, there is general agreement among linguists on its semantic contribution. Tense is a grammatical category that encodes time by locating eventualities (Lyons 1977: 678; Comrie 1985: 9; Huddleston & Pullum 2002: 116). In addition, tense locates eventualities relative to the utterance situation, that is, it is a deictic category (Comrie 1985: 14; Michaelis 2006: 220). Tense is generally considered to be an inflectional category. This raises issues about the future tense, since there is crosslinguistic evidence that future tenses are different from other tenses in that they may be expressed morphologically and periphrastically (Bybee & Dahl 1989: 90). Future time reference may be marked by morphological tenses in Romance languages and by modal auxiliaries in Germanic languages (see Celle 1997), alongside other periphrastic forms both in Romance and Germanic languages (see Lansari 2009).

For instance, Jespersen (1924: 50) states that French has a real future tense whereas English, which uses the modal auxiliary ‘will’ to refer to the future, has no such tense (see also Palmer 1974 and Huddleston & Pullum 2002: 208–209). As a modal auxiliary, ‘will’ may have uses that are unrelated to futurity (epistemic use: ‘That will be the postman’; habitual use: ‘She will sit there for hours doing nothing’; generic use: ‘Oil will float on water’). This formal difference has a semantic consequence, future time reference being potentially colored with a volitional overtone in English. Moreover, the speculative character of the future makes it different from the past. As pointed out by Comrie (1985: 43), the past has taken place and is beyond control, whereas the future can only be predicted and is less definite: “one might argue that while the difference between past and present is indeed one of tense, that between future on the one hand and past on the other should be treated as a difference of mood rather than one of tense”. Modal auxiliaries may be future-oriented and compete with future tense forms to refer to the future. In addition, future tenses may have modal as well as temporal uses. In terms of cross-linguistic historical evolution, Ultan (1978: 83) notes that future tenses “evolve from and develop into modal categories representative of varying degrees of uncertainty”. It seems to confirm that futurity cannot be conceptually separated from modality (see Lyons 1968: 56; Bybee, Perkins & Pagliuca 1994: Chapter 7).

This close link between futurity and modality is also found in the conditional. In Romance languages (with the exception of Romanian), conditionals serve to express ulteriority in the past as well as evidential and epistemic modal meanings: “*Le ‘conditionnel-temps’ est toujours également un ‘conditionnel-mode’*” (Bres, Azzopardi & Sarrazin 2012: 10). This confirms Bybee et al.’s (1994: 253–256) findings on the future, even if the “future in the past” was not considered in their study. As stressed by Bres et al. (2012: 10), modal meanings are derived from temporal meanings in Romance languages. Indeed, modality is omnipresent in language and others have aimed at capturing the different types and varieties of modalities exhibited in diverse languages (see contributions from various perspectives in Leiss & Abraham 2014).

1.2 Evidentiality and epistemicity

Part 2 addresses the link between evidentiality and epistemicity. Evidentiality is the encoding of information source. In some languages, it is obligatory for a speaker to indicate how they acquired knowledge (see Chafe & Nichols 1986; Willett 1988; Guentchéva 1996; De Haan 1999; Aikhenvald 2004, 2014; Diewald & Smirnova 2010). A distinction is generally made between direct information (be it visual, auditory or other sensory) and indirect information, which can be obtained through inference or hearsay (see Willett 1988: 57). And yet, there is little consensus on how to delimit and categorize evidentiality.

Several authors adopt a very restrictive approach to evidentiality. Thus, Bybee (1985: 184), Mithun (1986: 89) and Anderson (1986: 275) consider evidentiality to be a strictly grammatical phenomenon. Although Aikhenvald (2004: 24) acknowledges that “the distinction between inflectional and derivational categories is not clear-cut” and that information source may be expressed by non-grammatical elements, she limits evidentiality to grammaticalized evidentials in languages where it is obligatory to encode information source. The corollary of this is that she views evidentiality as a category “in its own right” (ibid: 7), in contrast to Palmer (1986: 54) who subsumes evidentiality under the heading of epistemic modality. According to Aikhenvald (2004: 4), marking information source “is not a function of truth or falsity”. She regards evidential uses of other non evidential morphemes as extensions (ibid: 105–152). A problem with this approach is that it is difficult to clearly distinguish between primary meaning and extended meaning. The evidential use of modal *wollen* and *sollen* in German is a case in point. Followed by a perfect main verb, these modals in the present tense are part of an unambiguously evidential construction, although their primary meaning is not evidential as illustrated in the following examples taken from Mortelmans (2000: 138, 139):

- (1) [...] *Kohl soll – verständlich – geschäumt haben.*
‘Kohl is said to have been foaming with anger, which is understandable’.
- (2) *Er will das gesehen haben.*
‘He claims to have seen it’.

With *wollen*, the reported speaker coincides with the subject and the quotative meaning is correlated with “a negative epistemic attitude of the speaker towards the proposition” (Mortelmans 2000: 139). This reveals the well-established connection between the perfect and evidentiality, the perfect functioning as an “evidential trigger” (Abraham 1998: 197): “Emphasizing a connection between an existing result and a completed past action leads to evidentials whose meaning is triggered by inference from the state secondary to the primary event”. This also suggests that the evidential meaning may well have become the core meaning, and that this construction can convey the speaker’s stance as well (see Celle 2006), that is, an epistemological communicative function. This positioning allows aligning with the addressee (Du Bois 2007).

Conversely, Aikhenvald (2004) argues that evidential categories may have extensions into other categories, such as the category of mirativity. According to Aikhenvald (2004: 209), mirativity encodes the “speaker’s ‘unprepared mind’, unexpected new information, and concomitant surprise”. Despite the fact that in many languages, mirativity exists as an extension of the evidential category, Aikhenvald views mirativity and evidentiality as separate categories, in line with DeLancey (2001). However, as in the case of evidentiality, mirativity cannot be limited to morphological marking. Constructional approaches seem to provide a more adequate treatment of mirativity and mirative extensions by taking into account the grammar-lexicon continuum and the gradience of grammatical categories (Celle & Tsangalidis 2017; Sahoo & Lemmens 2017).

The restrictive approach to categories is criticized by many linguists who claim that evidentiality cannot be conceptually separated from epistemic modality (e.g., Palmer 1986; van der Auwera & Plungian 1998; von Stechow & Gillies 2010). Even the linguists who uphold the position that evidentiality and epistemic modality are distinct categories (Diewald & Smirnova 2010; Faller 2002; Nuyts 2017; Mushin 2001) acknowledge that inferential evidentiality and epistemic necessity overlap. In addition, there is a correlation between the type of evidence and epistemic strength. Boye and Harder (2009: 27–28) argue that direct evidence implies a higher degree of certainty than indirect evidence.

Adopting a functional-cognitive perspective, Boye and Harder (2009: 17–20) claim that the grammaticalization process of verbs of saying, perception verbs, appearance verbs and attitude verbs supports the idea that the boundary between grammatically- and lexically-coded meaning is a porous one. Cross-linguistically,

such verbs may give rise either to evidential grammatical elements or to evidential sentence adverbials. Along that line of reasoning, adverbs such as ‘reportedly’, ‘allegedly’, ‘supposedly’, which are not considered by De Haan (2005) and Aikhenvald (2004) do qualify as evidential elements (Celle 2009) and should not be regarded as peripheral. Notionally, both lexical and grammatical elements can contribute to expressing evidentiality. Boye (2012: 2–3) further argues that evidentiality and epistemic modality are subsumed under a broader conceptual domain, namely the modal domain of epistemicity, which he defines as follows “a generalization over the notions of epistemic justification and epistemic support: the philosophers’ notion of *justificatory support* [...]”. This indicates that at the core of evidentiality is not only a source of information, but a subjective experience. Crucially, recent studies (e.g. Korotkova 2016) show that evidentials encode self-knowledge through indexicality, perception and introspection. Korotkova highlights the first-person cognitive process expressed by evidentials and treats them on a par with subjective expressions.

1.3 Aspect

The third part of the volume gathers four articles dealing with aspect and past temporality. In typology, aspect and tense are traditionally seen as two distinct categories, aspect being defined as the grammatical marking of “internal temporal constituency of a situation” (Comrie 1976: 6) and tense as the grammatical marking of “expression of location in time” (Comrie 1985: 9). However, as noted by Guentchéva (2016: 2–3), other approaches (see Binnick 1991; Klein 2009) question this clear-cut dichotomy and claim that aspect is inherently temporal from a semantic viewpoint, thus positing a broader spectrum of “tense-aspect” marking systems. Comrie’s classical definition has also been nuanced by researchers who more particularly associate aspect with speaker viewpoint on the event (see e.g. Bertinetto & Delfitto 2000; Johanson 2000; Klein 2009; Smith 1991). This conception of aspect as speaker perspective is particularly well-represented in the French enunciative tradition, where the difference between the French tenses *imparfait* and *passé simple* is often encapsulated in terms of subjective internal viewpoint vs. objective external viewpoint (see Bres 2003 for an overview). Whereas the tense/aspect interface has been widely investigated, other cross-linguistic works shed light on lesser known phenomena, such as the intricate links between aspect, syntax and modality (see various contributions in Abraham & Kulikov 1999 and Abraham & Leiss 2008).

The lack of consensus regarding the definition of aspect may be linked to at least three elements. First, it must be borne in mind that typological models of aspect were originally strongly influenced by aspect marking in Slavic languages, where a fundamental aspectual opposition (imperfective vs. perfective) is encoded

at the lexical and derivational level. Cross-linguistically, the use of the labels ‘imperfective’ and ‘perfective’ to describe non-Slavic systems, where aspect may be encoded by morphemes for example, has proven difficult and has even obscured the discussion on aspect and the terminology used (see Guentchéva 2016: 3). The theoretical issue raised here is inherent to typology: comparing languages calls for the identification of stable conceptual or semantic categories, but the mapping of such categories onto very diverse language systems may prove complicated (see LaPolla 2016 for a full-fledged discussion on categorization). Recent research (see the CAMNAM Project¹) actually denounces the “indo-euro-centrism” that has infused the linguistic conceptualization of tense and aspect. Confronting the general typological category of aspect – and its many sub-categories (imperfective vs. perfective, bounded vs. unbounded, etc.) devised by linguists such as Vendler’s (1967) – with lesser studied non-Indo-European languages is a promising path for future research on aspect.

Secondly, the lack of consensus regarding aspect is tightly linked to categorization processes and what we mean by ‘category’ in linguistics: is aspect a conceptual/semantic category, in which case it can remain covert and be merely expressed by verbal semantics in some languages? Or is aspect necessarily marked at the grammatical level? The former view on aspect is usually referred to as *Aktionsart* or lexical aspect, but many theoretical frameworks recognize that lexical aspect and grammatical aspect strongly interact (see Guentchéva 2016: 6). These various interactions between grammatical aspect, verbal semantics and various complements (verb objects but also adverbial modifiers of the type ‘for one day’/ ‘in ten minutes’) are now referred to as “aspectual coercion effects” (see De Swart 2000). There is therefore a need to posit various degrees of tense-aspect grammaticalization depending on the languages at stake: in some languages, tense-aspect is encoded through lexical means, while in others it may be more grammaticalized. This cline of grammaticalization might prove a decisive parameter in cross-linguistic investigations of aspect.

Thirdly, the variety of approaches and theoretical frameworks found in the literature on aspect may have made cross-linguistic comparisons of aspect marking particularly uneasy. Typologists tend to adopt an onomasiological perspective on aspect, as they seek to find linguistic realizations of conceptual categories (see for instance the research carried out within the EUROTYP group by Dahl (2000), who assumes the cross-linguistic existence of a “progressive gram”), while descriptive linguists focus on specific markers within a given language system,

1. The CAMNAM project is an interdisciplinary project bearing on Khmer languages and cultures <<https://www.vjf.cnrs.fr/sedy1/recherches.php?voirlong=O&type=projet&programme=CAMNAM>>.

from a semasiological viewpoint. Language-specific descriptive research may actually enable us to develop new ways of looking at aspect marking. For instance, recent innovative research in French tries to relate semantic analyses of aspect to computational models (see Gosselin et al. 2013). Another promising current trend comes from cognitive linguistics (see i.e. De Wit 2017 on the “present perfective” from a cross-linguistic viewpoint), and more particularly from Construction Grammar, a usage-based cognitive theory that posits that the basic linguistic units are “constructions”, defined as conventionalized form-meaning pairings (see Goldberg 2006). Within this theory, several works on aspectual constructions in English (Michaelis 1998, 2004; Croft 2012) show that aspect may be entrenched in conventionalized lexico-grammatical constructions that imply discourse-pragmatic factors.

2. Presentation of the volume

The present volume sheds new light on TAM-E issues, both theoretically and empirically. From a theoretical point of view, it reasserts well-known universal tendencies – such as the tendency for tense markers to develop modal meanings and for modal markers to acquire “post-modal” usages, or the universality of the “perfect” category – while offering fine-grained analyses pointing to cross-linguistic differences that had gone unnoticed so far (see the contributions by Hütsch, Treikelder & Amon, Verhees, Corre).

As a whole, the volume also strongly reaffirms the relevance of the semantic categories of evidentiality and mirativity for languages that do not encode these categories in their grammatical systems (see for instance the contributions by Marín Arrese, Carretero & Berdasco-Gancedo, Celle on Spanish and English). These semantic categories are used to account for a wide variety of phenomena such as modal verbs or evidential expressions, but also specific types of questions (in Celle), non-finite constructions (in Usoniené & Vincent) or discourse markers (in Tan & Mursell, Ranger). The apparent heterogeneity of the markers under scrutiny raises major categorization issues: what are the criteria used to classify a specific marker as modal, evidential or mirative? On the basis of the deictic character of evidentials, a connection is established between evidentials and expressives (see Celle). Conversely, non-finite forms being unrelated to the utterance situation, their evidential contribution relies on the governing verb (see Usoniené & Vincent). In other words, what is at stake is how language-specifics interact with broad semantic cross-linguistic categories.

The present contributions offer different answers depending on the theoretical framework adopted, which opens up challenging discussions for further research.

Some of them offer diachronic accounts, either confirming well-entrenched grammaticalization paths (chapters by Senti, Bres et al.), or pointing to diachronic innovations in the case of split systems (see Usonienė & Vincent).

From a more empirical perspective, several chapters are strong descriptive contributions focusing on linguistic phenomena that had not been investigated so far or rather poorly examined. Some contributions center around languages that had been largely overlooked in the literature on TAM-E (Old Catalan in Senti, Lithuanian in Usonienė & Vincent, East Caucasian languages Avar and Andi in Verhees). Others focus on supposedly well-known markers – the periphrastic future-in-the-past in French in Bres et al., the comparison between *passé composé* and *imparfait* in French as well in Wicher, motion verbs in European Portuguese in Leal et al. – but they do so by relying on new corpus-based data and quantitative approaches (for Bres et al. and Wicher) or/and test them against theoretical frameworks (for Leal et al. and Wicher) to offer innovative ways to account for the markers under scrutiny.

We now present each of the chapters of the volume individually. Each chapter is summarized in order to provide its theoretical framework and data as well as the main results of the analyses carried out.

2.1 Part I: Futurity, modality, conditionals

Chapter 2 by Annalena Hütsch (University of Neuchâtel) – entitled ‘A quantitative perspective on modal values of future tense in French and German’ – proposes to analyze the modal values being conveyed by the future tense in French and German by using comparable corpora of newspaper texts (*Le Monde* 2007, *Süddeutsche Zeitung* 2013), from both a qualitative and a quantitative perspective as well as within an enunciative approach. The theoretical framework adopted is called “triadic model of analysis” established by Rossari (2016) which provides a new concept of how modal forms operate. The differences of modal uses are considered for *futur simple* and *futur antérieur* for French vs. *Futur I* and *Futur II* for German. Besides the fact that modal uses of future tense are rather rare in both languages compared to their temporal use, the quantitative analysis shows that French appears to have a more modal use of future tense than German. The future tense has a greater proportion of modal rather than temporal meanings in French than in German, with higher relative frequencies of modal occurrences and a greater range of modal values in the newspaper text genre.

Chapter 3, co-authored by Anu Treikelder and Marri Amon (University of Tartu), bears on ‘The temporal uses of modal verbs: A comparison of French *devoir* and Estonian *pidama*’. The authors rely on a translation corpus to compare the temporal uses of two modal verbs, French *devoir* and Estonian *pidama* in their

reference to the future. While they share similar properties in root modality uses, these verbs present considerable differences in their epistemic and postmodal uses. *Devoir* has been labelled “auxiliary of the future” in the process of grammaticalization, while *pidama* has a more limited temporal value: out of its three postmodal values – quotative, avertive and intentional – only the latter has been associated with the future. The analysis of translations indicates that *pidama* does not have the most grammaticalized temporal use, that is, the “objective future in the past” of the verb *devoir*, while revealing a rather high correspondence between the two verbs in the other future-related uses reported for *devoir*. The authors examine different situation types that occur in temporal uses and discuss the interaction of modal meanings with future time reference not only to specify the temporal meanings in each language, but also to contribute to the understanding of the temporal reference of modal verbs in general.

Chapter 4, entitled ‘The competition between present conditional and prospective imperfect in French: a diachronic analysis’, is co-authored by Jacques Bres, Sascha Diwersy and Giancarlo Luxardo (Université Paul Valéry, Montpellier). Extensive work has been carried out about the competition faced over the centuries by the synthetic future (*partira*) compared to the periphrastic turn with a prospective present (*va partir*). Much less attention has been paid to the parallel competition between the present conditional (*partirait*) and the periphrastic turn with prospective imperfect (*allait partir*). It is the object of the present study, limited to the competition between the two forms in their tense usages. The authors first make explicit the usages in discourse of each form according to their diachronic appearance: subjective ulteriority (10th century) and objective ulteriority (19th century) for the conditional; imminence (15th century), subjective ulteriority (16th century) and objective ulteriority (late 18th century) for the prospective imperfect. The two forms compete with each other in the expression of ulteriority, both subjective and objective. Based on PRESTO, a multigenre diachronic corpus of 16th–20th century annotated texts, the study only addresses the question of diachronic evolution of the ratio present conditional/ prospective imperfect in the expression of subjective ulteriority. The methodology used (mainly: text retrieval and hierarchical clustering) shows that the prospective imperfect is steadily increasing.

2.2 Part II: Evidentiality and epistemicity

Chapter 5 by Juana I. Marín Arrese (Universidad Complutense de Madrid) deals with ‘Evidentiality and the TAM systems in English and Spanish: a cognitive and cross-linguistic perspective’. This chapter presents results of a contrastive case study (English vs. Spanish) on TAM markers expressing evidential functions: (a) epistemic

modals ‘must’, ‘must have’, and *deber* (*de*), *deber* (*de*) *haber*; (b) future perfect verbal forms ‘will have’, ‘*habrá/n*’; and (c) conditional perfect verbal forms ‘would have’, ‘*habría/n*’. The aim is to identify and describe the main attested extensions of these TAM markers to inferential and reportative values of evidentiality. The author also aims to explore the degree to which these evidential values may be associated with particular discourse domains (written journalistic vs. unscripted conversation). The data consist of naturally occurring examples, randomly selected from spoken and written corpora in the two languages (BNC-Baby, CORLEC-UAM, CESJD-UCM). It is argued that certain parameters play a crucial role in facilitating these extensions: immediacy in relation to the ground, immediate vs. non-immediate, and reality, real vs. projected are mainly involved in inferential values, and the feature irrealis and frame-shifting to an alternative point of view in the case of reportative evidentiality.

Chapter 6, co-authored by Marta Carretero (Universidad Complutense de Madrid) and Yolanda Berdasco-Gancedo (Madrid Open University), also compares English and Spanish as far as evidentiality is concerned. This chapter is entitled ‘Evidentiality in English and Spanish informative financial texts: a contrastive analysis’ and sets forth a contrastive analysis of evidential expressions in English and Spanish financial texts. Apart from standard evidential expressions, the analysis also includes evidential frames. The choice of informative financial texts was motivated by the importance of evidential expressions in this register, in which journalists often resort to sources of evidence for drawing conclusions or making predictions about financial matters. The corpus consists of 80 texts written in British English and Peninsular Spanish, 40 from specialized financial journals and 40 from non-specialized quality journals, totalling about 50,000 words. The texts have been selected according to the following criteria: financial and macro-economic issues as topics, length of at least 500 words and publication between 2013 and 2015. All the occurrences of evidentials in the texts were analyzed and submitted to a quantitative analysis. The results show the predominance of communicative, highly reliable and intersubjective evidentiality: not surprisingly, journalists rely mainly on linguistic messages as a source of evidence and present evidential qualifications as creditable and shareable by others. This pattern is common across languages and levels of specialization. However, these factors account for differences in frequency and distribution of evidentials.

Chapter 7 authored by Andreu Sentí (University of València) is a diachronic study devoted to ‘Evidentiality and epistemic modality in Old Catalan. A diachronic cognitive approach to the semantics of modal verbs’. Epistemic modality and evidentiality are two categories that have not been clearly defined in the literature as in the case of the Romance verbs *deure*, *devoir*, *dovere*, *deber* (‘must’) and *poder*, *pouvoir*, *potere* (‘may’). In order to clarify the relationship between these two

domains, the author draws a detailed semantic map for Catalan modal verbs using a corpus-based study of modal verbs *deure* ('must'), *haver de* ('have to') and *poder* ('can/may') in Old Catalan (11th–16th centuries). He argues that the innovative readings of the modal verbs *deure* and *haver de* are evidential because they primarily encode the source of information. If these verbs allow an epistemic reading, it is a pragmatic contextual value, but it is not related to the meaning encoded by the verbs. On the other hand, it is hypothesized that *poder* encodes epistemic values. Corpus data show a diachronic tendency towards more subjective values. The semasiological change of the modal verb *deure* involves the development and the strengthening of a more subjective inferential value, a generic inference. However, *haver de* develops a more objective inferential value, a specific inference. In the case of *poder*, the innovative value is also subjective, but there are no explicit premises in the context. In this case, the speaker makes a hypothesis, and the epistemic commitment is downgraded. Therefore, the subjective values should not be confused with the higher or lower degrees of certainty (i.e. epistemic modality).

Chapter 8 by Graham Ranger is entitled 'An enunciative and corpus-based perspective on 'I think' and investigates epistemicity and evidentiality in relation to discourse marking. This chapter more particularly focuses on 'I think' as a discourse marker, used to indicate stance or epistemic modality; It is well known that 'I think' may assume a variety of different functions such as shielding, approximator, structural or booster which may reflect specific configurations dependent on identifiable contextual features, but the author first proposes to test the hypothesis that 'I think' is not inherently ambiguous with a corpus-based investigation of collocational affinities of the sequence, which will reveal a number of characteristic environments (British National Corpus). Then, the author suggests an enunciative description of 'I think' in terms of a basic schematic form, which undergoes certain controlled and calculable deformations to generate local 'shapes', that is, contextually situated values. The theoretical objectives of the analysis are to show that: (1) if a marker is not contextually situated, then it can only be described in terms of its potential for meaning; (2) the value of a marker depends both on the surrounding context (including prosody), and on its own specific latitudes of deformability. The author concludes that 'I think' in itself expresses neither evidentiality nor epistemic modality, but that it is compatible with both because of its contextual configurations.

Chapter 9 by Jennifer Tan (Autonomous University of Madrid) and Johannes Mursell (Goethe-Universität Frankfurt), entitled 'Embedding evidence: on two types of evidentials', tackles the issue of evidentiality in relation to discourse particles and epistemic modals. Crosslinguistically, evidentials differ along certain parameters, such as embeddability. The authors argue that German *wohl* and Tagalog *yata* should be considered as evidentials because they indicate both source and reliability of information, since they allow the speaker to express that s/he has

some piece of evidence to believe the propositional content of the utterance, but is uncertain about it. However, their distribution in embedded clauses sets them apart. Contrasts in embedding have led to a treatment of evidentials as either illocutionary modifiers or epistemic modals. Using natural occurrences and a corpus of parliament speeches, the authors show that *wohl* can only be embedded in a subset of the contexts in which embedding *yata* is possible. This is expected since *wohl*, as discourse particle or illocutionary modifier, depends on the presence of illocutionary force, entering a syntactic agreement relation with it (Coniglio & Zegrean 2012), while *yata*, as an epistemic modal, does not. It is thus predicted that *wohl* is found in embedded clauses containing illocutionary force while the distribution of *yata* is less restricted. The analysis has two important consequences: (a) evidential markers need a non-uniform treatment as either illocutionary modifiers or epistemic modals; (b) constraints on embeddability are a valid diagnostic to distinguish these types, as the speech act operator type of evidentials is dependent on illocutionary force and is thus much more restricted in its distribution in embedded clauses than the less limited epistemic modals.

Chapter 10 by Agnès Celle (Université Paris Diderot and University of Colorado) deals with mirativity and speech acts, and focuses on ‘Questions as indirect speech acts in surprise contexts’. This chapter offers an analysis of two types of interrogatives used as indirect speech acts in surprise contexts in English – so-called unresolvable questions (‘What the hell is this?!’) and rhetorical questions (‘How can you just walk around like that?’). The function of these questions is not to request information that is unknown to the speaker. It is argued that surprise-induced unresolvable questions are expressive speech acts devoid of epistemic goals. This suggests that expressives serve an evidential function in interrogatives. Surprise-induced rhetorical questions are shown not to propose an obvious answer, but to request a commitment update from the addressee. This leads to integrating the commitment update parameter into a typology of questions. Adopting a schema-theoretic approach to surprise, it is shown that unresolvable questions and rhetorical questions are linguistic expressions of mirativity, the former manifesting the initial stage of the cognitive processing of unexpectedness, the latter the last stage.

Chapter 11, co-authored by Aurelija Usonienė (Vilnius University) and Nigel Vincent (University of Manchester) is entitled ‘Non-finiteness and evidentiality: the Lithuanian *Accusativus cum Participio* in a cross-linguistic perspective. The fact that non-finite constructions and evidentiality are related has been observed in various European languages. The use of passive matrix verbs with the infinitive in English and the corresponding though much less productive pattern in Dutch, reportive passives in Danish, and evidential participle constructions in Lithuanian have all received much attention in the literature. This chapter is devoted to the less studied grammatical realization of indirect evidentiality, namely the construction

traditionally called *Accusativus cum Participio*. In contemporary Lithuanian, it is only found in the complementation of communication, cognition and perception verbs. The purpose of the analysis is to show that the use of non-finite non-agreeing participial *be*-verb forms is obligatory for the complement phrase to have a propositional status. In this respect the use of the non-agreeing participle can be compared to the obligatory use of the full infinitive *to be* with raising verbs in English (*He seems to be away; I believe him to be happy*). The quantitative and qualitative corpus-based analysis is used to investigate the distribution of the construction in different types of discourse (fiction, academic and journalistic). The analysis is set in the framework of a typological comparison of similar uses of non-finite verb forms in other European languages. It shows that non-finite forms correlate with evidentiality. It also points to a diachronically-innovative feature of Lithuanian with respect to agreement, as this language exhibits a diachronic split in the marking of non-finite forms.

2.3 Part III: Aspect and past temporality

Chapter 12 by Samira Verhees (National Research University, Moscow) is entitled ‘How perfect are East Caucasian perfects? A corpus approach to grammatical semantics in Avar and Andi’ questions the aspectual category of ‘perfect’. This chapter deals more particularly with the semantics of the perfect in Avar and Andi, two East Caucasian languages in which the perfect forms belong to a type commonly found in the larger geographical area where they are spoken. They are formed with a ‘be’-auxiliary and feature a prominent expression of indirect evidential semantics, alongside other meanings such as resultative. Descriptive literature suggests that the lexical semantics of the verb play an important role in the interpretation of the perfect, but no quantitative research currently exists on which lexical items actually enter in this construction more frequently in language use, and consequently, which meanings of this polysemous form result from specific combinations. By providing a clear-cut classification of the possible meanings these forms can have cross-linguistically, the patterns of their usage in particular languages can be described more adequately, as is shown in this study with data from elicitation and two corpora. The presence of an agent in the ergative is shown to be an important parameter in distinguishing resultatives proper from resultative perfects in these languages. This distinction is relevant to determine whether current relevance meanings of the perfect are at all represented in these languages, as they are not very well-attested. This study shows that the Avar perfect represents a highly polysemous verb form that combines resultative proper, current relevance and indirect evidentiality, while Andi shows a more advanced stage of grammaticalization of the indirect evidential meaning.

Chapter 13 by Eric Corre (Université Paris 3-Sorbonne Nouvelle) is devoted to ‘The different grammars of event singularization: a cross-linguistic corpus study’. It proposes a comparative approach to the expression of bounded single situations or episodes across four languages, based on *L’Etranger* (*The Outsider*) by Albert Camus and three published translations (English, Russian, Hungarian). *L’Etranger* is notorious for its use of the French tense-aspect *p(assé) c(omposé)*, a primary candidate for marking boundedness. In a binary approach to aspect such as Smith’s (1991) two-component theory, situation aspect combines with viewpoint aspect to compute the aspectual composition of sentences. Other, typologically-oriented models, have shown that the universal category aspect need not look like Slavic aspect, because what we find across languages are polysemous aspectual clusters, heterogeneous ‘grams’. One goal of this chapter is to determine, based on usage, which system is more descriptively adequate. The PC often yields bounded, quasi-telic meanings in and of itself; in those cases, English marks perfectivity more explicitly on the DP-level or otherwise, as the simple past is aspectually neutral; it is shown that the SP comes out as ambiguous, yielding perfective as well as imperfective readings. Russian and Hungarian have a different strategy: they rely heavily on derivational, classificatory prefixes to turn an unbounded situation into a (de)limited one. The prefixed verbs’ bias towards singularized situations explains the absence of limitative perfectivity for atelic predicates of the French or English type, expressed by bare verbs in Russian and Hungarian. However, the two languages diverge notably: the former has morphologized aspect, not the latter; a Hungarian unprefix stem may denote a bounded or unbounded situation, especially in certain classes of predicates (consumption, creation). This chapter shows that only an empirically-grounded, fine-grained analysis in the expression of event singularisation, can shed light on the nature of aspectual systems.

Chapter 14 by Oliver Wicher (Paderborn University) bears on the comparison between two French tenses. It is entitled ‘A corpus-driven account of French past tenses: *passé composé* and *imparfait* revisited’ and presents a corpus-driven analysis of usage patterns of the French past tenses *passé composé* (PC) and *imparfait* (IMP). Conventional wisdom has it that these two tenses encode perfective and imperfective aspect, respectively, indicating that past-tense choice is considered a matter of the speaker’s perspective. Although a vast amount of research on French past tenses has analyzed related topics such as aspectual compositionality, event structure or discourse grounding, only a few studies, have focused on distributions of French past tenses in attested data and they suffer from the following caveats: (a) they are based on small-sized corpora of specific genres such as sports commentaries, newspapers or obituaries; (b) they are confined to giving raw or percentage-wise frequencies of past-tense occurrences in text, neglecting construction grammar approaches. To fill this gap, the author introduces a new genre-diverse reference corpus

of French, the *Corpus de référence du français contemporain* (CRFC) and puts it to the test. Assuming that past-tense choice correlates with the lexico-environment of the verb, this chapter uses inferential statistics to calculate the collostructional strength. Adopting a construction grammar perspective, the author proposes to analyze past-tense constructions of high-frequent polysemous verbs to test our main assumption that the distinction between PC and IMP can be considered an alternation phenomenon. It is further hypothesized that characteristic past-tense collostructions can be revealed, differing in terms of their preferred complements. The results show that a corpus-driven and constructional approach provides new insights into the usage of French past tenses. Past tense occurrences of *vouloir* are shown to predominantly collocate with verbal complements. In the imparfait, *vouloir* tends to occur in tentative constructions, while perfective *vouloir* correlates with dynamic verbal phrases. The past tense of *voir* is more frequent with perfective than imperfective constructions. This chapter demonstrates how a corpus-driven lexico-grammatical approach to French past-tensed verbs can help explore their usage patterns. This approach opens up avenues for future research into the acquisition of past tenses in SLA.

Chapter 15 by António Leal, Fátima Oliveira and Purificação Silvano focuses on aspect in relation to motion verbs. The main objective of this chapter, entitled “Aspectual issues on verbs of inherently directed motion and goal prepositions in European Portuguese”, is to put forward a proposal that accounts for the semantics of inherently directed motion verbs *ir* (‘go’) and *vir* (‘come’) combined with prepositional phrases (PPs) with the thematic role of Goal, headed by the prepositions *para* (‘to/towards’) and *até* (‘to’) in European Portuguese. Based on the notion of path scale, the chapter assumes that, like degree achievements, motion verbs also have an underlying scalar structure, which is underspecified in the following way: the measuring dimension is specified (paths); the set of degrees, namely if there is a maximum value in the scale, and the ordering relation (proximity, or not, to a certain point) are not specified. It is argued that these PPs contribute in a distinct way to the specification of one of the parameters of path scales: *para* specifies the ordering relation (approach to a certain point defined by the *para*-PP), whereas *até* operates on the parameter of set of degrees and denotes a maximum element that is contextually relevant, and, thus, transforms the scale projected by the verb in a closed scale. Furthermore, the authors assume that verbs *ir* and *vir* are underspecified, that is, they allow the contribution of other elements of the predication or the context if the verb is associated with a two point scale or to a multi-point scale. In our proposal, this is performed by PPs such as *para* and *até*, whose different semantics give rise to different predication readings with *ir* and *vir*. Scale semantics is shown to provide an adequate framework to account for the underlying scalar structure of inherently directed motion verbs.

References

- Abraham, Werner. 1998. The morphological and semantic classification of 'evidentials' and modal verbs in West Germanic: The perfect(ive) catalyst. *Groninger Arbeiten zur germanistischen Linguistik* 42: 192–206.
- Abraham, Werner & Kulikov, Leonid (eds). 1999. *Tense-aspect, Transitivity and Causativity. Essays in honor of Vladimir Nedjalkov* [Studies in Language Companion Series 50]. Amsterdam: John Benjamins. doi:10.1075/slcs.50
- Abraham, Werner & Leiss, Elisabeth (eds). 2008. *Modality-Aspect Interfaces: Implications and Typological Solutions* [Typological Studies in Language 79]. Amsterdam: John Benjamins. doi:10.1075/tsl.79
- Aikhenvald, Alexandra. 2004. *Evidentiality*. Oxford: OUP.
- Aikhenvald, Alexandra. 2014. The grammar of knowledge: A cross-linguistic view of evidentials and the expression of information source. In *The Grammar of Knowledge. A Cross-Linguistic Typology*, Alexandra Aikhenvald & Robert M. W. Dixon (eds), 1–51. Oxford: OUP. doi:10.1093/acprof:oso/9780198701316.003.0001
- Anderson, Lloyd. 1986. Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In *Evidentiality: The Linguistic Coding of Epistemology*, Wallace Chafe & Johanna Nichols (eds), 273–312. Norwood NJ: Ablex.
- Bertinetto, Pier Marco & Delfitto, Denis. 2000. Aspect vs. actionality: Why they should be kept apart. In *Tense and Aspect in the Languages of Europe*, Östen Dahl (ed.), 189–225. Berlin: Mouton de Gruyter.
- Binnick, Robert, I. 1991. *Time and the Verb: A Guide to Tense and Aspect*. Oxford: OUP.
- Boye, Kasper. 2012. *Epistemic Meaning: A Crosslinguistic and Functional-cognitive Study*. Berlin: Mouton de Gruyter. doi:10.1515/9783110219036
- Boye Kasper & Harder, Peter. 2009. Evidentiality: Linguistics categories and grammaticalization. *Functions of Language* 16: 9–43 doi:10.1075/fol.16.1.o3boy
- Bres, Jacques. 2003. Temps verbal, aspect et point de vue: De la langue au discours. *Cahiers de Praxématique* 41: 55–84.
- Bres, Jacques, Azzopardi, Sophie & Sarrazin, Sophie. 2012. *Présentation générale: Ulteriorité du passé et valeurs modales. Faits de Langues* 40: 5–11.
- Bybee, Joan, Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar*. Chicago IL: University of Chicago Press.
- Bybee, Joan & Dahl, Östen. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13(1): 51–103. doi:10.1075/sl.13.1.o3byb
- Bybee Joan. 1985. *Morphology. A Study of the Relation between Meaning and Form* [Typological Studies in Language 9]. Amsterdam: John Benjamins. doi:10.1075/tsl.9
- Chafe, Wallace & Nichols, Johanna (eds). 1986. *Evidentiality: The Linguistic Coding of Epistemology*. Norwood NJ: Ablex.
- Celle, Agnès. 1997. *Etude contrastive du futur français et de ses réalisations en anglais*. Paris: Ophrys.
- Celle, Agnès. 2006. *Temps et modalité: l'anglais, le français et l'allemand en contraste*. Bern: Peter Lang.
- Celle, Agnès. 2009. Hearsay adverbs and modality. In *Modality in English: Theory and Description*, Raphael Salkie, Pierre Busuttill & Johan van der Auwera (eds), 269–293. Berlin: Mouton de Gruyter. doi:10.1515/9783110213331.269

- Celle, Agnès & Tsangalidis, Anastasios. 2017. Introduction. *Review of Cognitive Linguistics, The Linguistic Expression of Mirativity* 15(2): 305–311.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: CUP.
- Comrie, Bernard. 1985. *Tense*. Cambridge: CUP. doi:10.1017/CBO9781139165815
- Croft, William. 2012. *Verbs: Aspect and Causal Structure*. Oxford: OUP.
doi:10.1093/acprof:oso/9780199248582.001.0001
- Dahl, Östen. 2000. The tense-aspect systems of European languages in a typological perspective. In *Tense and Aspect in the Languages of Europe*, Östen Dahl (ed.), 3–25. Berlin: Mouton de Gruyter.
- de Haan, Ferdinand. 1999. Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics* 18: 83–101.
- de Haan, Ferdinand. 2005. Semantic distinctions of evidentiality. In *The World Atlas of Language Structures*, Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie (eds), 314–317. Oxford: OUP.
- DeLancey, Scott. 2001. The mirative and evidentiality. *Journal of Pragmatics* 33: 369–382.
doi:10.1016/S0378-2166(01)80001-1
- de Swart, Henriëtte. 2000. Tense, aspect and coercion in a cross-linguistic perspective. In *Proceedings of the Berkeley Formal Grammar Conference*, Miriam Butt & Tracy Holloway (eds), 1–20. Stanford CA: CSLI.
- De Wit, Astrid. 2017. *The Present Perfective Paradox across Languages*. Oxford: OUP.
- Diewald, Gabriele & Smirnova, Elena. 2010. *Linguistic Realization of Evidentiality in European Languages*. Berlin: Mouton de Gruyter. doi:10.1515/9783110223972
- Du Bois, John. 2007. The stance triangle. In *Stancetaking in Discourse* [Pragmatics & Beyond New Series 164], Robert Englebretson (ed.), 139–182. Amsterdam: John Benjamins.
doi:10.1075/pbns.164.07du
- Faller, Martina. 2002. *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. PhD dissertation, Stanford University.
- Fleischman, Suzanne. 1982. *The Future in Thought and Language: Diachronic Evidence from Romance*. Cambridge: CUP.
- Gosselin, Laurent, Enjalbert, Patrice, Mathet, Yann & Becher, Gérard. 2013. *Aspect de l'itération. L'expression de la répétition en français: Analyse linguistique et formalisation*. Bern: Peter Lang.
doi:10.3726/978-3-0352-0215-1
- Goldberg, Adele. 2006. *Constructions at Work*. Oxford: OUP.
- Guentchéva Zlatka (ed.). 1996. *L'énonciation médiatisée*. Louvain: Peeters.
- Guentchéva, Zlatka. 2016. Introduction. In *Aspectuality and Temporality: Descriptive and Theoretical Issues* [Studies in Language Companion Series 172], Zlatka Guentchéva (ed.), 1–24. Amsterdam: John Benjamins.
- Huddleston, Rodney & Pullum, Geoffrey. 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Jespersen, Otto. 1924. *The Philosophy of Grammar*. London: Allen & Unwin.
- Johanson, Lars. 2000. Viewpoint operators in European Languages. In *Tense and Aspect in the Languages of Europe*, Östen Dahl (ed.), 27–187. Berlin: Mouton de Gruyter.
- Klein, Wolfgang. 2009. How time is encoded. In *The Expression of Time*, Wolfgang Klein & Ping Li (eds), 39–82. Berlin: Mouton de Gruyter. doi:10.1515/9783110199031.39
- Korotkova, Natalia. 2016. *Heterogeneity and Uniformity in the Evidential Domain*. PhD dissertation, University of California at Los Angeles.
- Lansari, Laure. 2009. *Les périphrases verbales aller + infinitif et be going to*. Paris: Ophrys.

- LaPolla, Randy J. Forthcoming. On categorization: Stick to the facts of the languages. *Linguistic Typology*.
- Leiss, Elisabeth & Abraham, Werner (eds). 2014. *Modes of Modality. Modality, Typology, and Universal Grammar* [Studies in Language Companion Series 149]. Amsterdam: John Benjamins. doi:10.1075/slcs.149
- Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: CUP. doi:10.1017/CBO9781139165570
- Lyons, John. 1977. *Semantics*. Cambridge: CUP.
- Manning, Liliann, Cassel, Daniel & Cassel, Jean-Christophe. 2013. St Augustine's reflections on memory and time and the current concept of subjective time in mental time travel. *Behavioral Sciences* 3(2): 232–243. doi:10.3390/bs3020232
- Michaelis, Laura A. 1998. *Aspectual Grammar and Past-Time Reference*. London: Routledge.
- Michaelis, Laura A. 2004. Type shifting in Construction Grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15: 1–67. doi:10.1515/cogl.2004.001
- Michaelis, Laura. 2006. Tense in English. In *The Handbook of English Linguistics*, Bas Aarts & April McMahon (eds), 220–243. Oxford: Blackwell. doi:10.1002/9780470753002.ch10
- Mithun, Marianne. 1986. *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. PhD dissertation, Stanford University.
- Mortelmans, Tanya. 2000. On the 'evidential' nature of the 'epistemic' use of the German modals *müssen* and *sollen*. *Belgian Journal of Linguistics* 14: 131–148. doi:10.1075/bjl.14.08mor
- Mushin, Ilana. 2001. *Evidentiality and Epistemological Stance: Narrative Retelling* [Pragmatics & Beyond New Series 87]. Amsterdam: John Benjamins. doi:10.1075/pbns.87
- Nuyts, Jan. 2017. Evidentiality reconsidered. In *Evidentiality Revisited* [Pragmatics & Beyond New Series 271], Juana Isabel Marin Arrese, Gerda Haßler & Marta Carretero (eds), 57–83. Amsterdam: John Benjamins. doi:10.1075/pbns.271.03nuy
- Palmer, Frank R. 1986. *Mood and Modality*. Cambridge: CUP.
- Poutsma, Hendrik. 1922. *Mood and Tense of the English Verb*. Groningen: P. Noordhoff.
- Prior, Arthur. 1967. *Past, Present and Future*. Oxford: Clarendon Press. doi:10.1093/acprof:oso/9780198243113.001.0001
- Sahoo, Kalyanamalini & Lemmens, Maarten. 2017. Degrees of mirativity. *Review of Cognitive Linguistics, The Linguistic Expression of Mirativity* 15(2): 343–384.
- Smith, Carlota. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer. doi:10.1007/978-94-015-7911-7
- Ultan, Russell. 1978. The nature of future tenses. In *Universals of Human Language, IV: Word Structure*, Joseph Greenberg (ed.), 83–123. Stanford CA: Stanford University Press.
- van der Auwera, Johan & Plungian, Vladimir. 1998. Modality's semantic map. *Linguistic Typology* 2: 79–124. doi:10.1515/lity.1998.2.1.79
- van Wassenhove, Virginie. 2017. Time consciousness in a computational mind/brain. *Journal of Consciousness Studies* 24(3–4): 177–202.
- Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca NY: Cornell University Press.
- von Stechow, Kai & Gillies, Anthony S. 2010. Must ... stay ... strong! *Natural Language Semantics* 18(4): 351–383. doi:10.1007/s11050-010-9058-2
- Willett, Thomas. 1988. A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language* 12(1): 51–97. doi:10.1075/sl.12.1.04wil

A quantitative perspective on modality and future tense in French and German

Annalena Hütsch

Université de Neuchâtel

This chapter looks at modal nuances conveyed by future tenses in French (*futur simple*, *futur antérieur*) and German (*Futur I*, *Futur II*) via a corpus-based study, using comparable newspaper corpora (*Le Monde*, *Süddeutsche Zeitung*). In addition to a qualitative analysis based on an enunciative approach to modal forms (Rossari 2016), we will adopt a quantitative perspective in order to elicit statistical evidence on the nature and degree of modality expressed by the future tense in daily newspapers. Besides the fact that modal use of future tenses is rather rare in both languages compared to their temporal use, the quantitative analysis shows that French appears to have a more modal use of future tense than German in the text type sampled.

Keywords: modality, future tense, French, German, corpus study, quantitative analysis

1. Introduction

Modal meaning of future tense has already been the focus of numerous studies, whether the target language was French, German or both (Celle 2006; Confais 1992; Fleischman 1982; Fritz 2000; Revaz 2009; Sundell 2003; Vater 1975, to name but a few).¹ This chapter proposes to look at modal nuances conveyed by future tenses in French and German in a different light. Using comparable corpora of newspaper texts, we will adopt a qualitative as well as a quantitative perspective in order to elicit statistical evidence on the modal nature of future tenses in daily newspapers. Quantitative analyses of this topic are rare, and almost inexistent regarding comparisons between French and German. We hope the present study will start to fill this critical gap. The paradigm under consideration is the *futur simple* and the *futur*

1. Cf. also Damourette & Pichon (1911–1936: § 1820ff.), whose grammar paved the way for subsequent studies on temporal and modal future tense uses in French.

antérieur for French vs. the *Futur I* and the *Futur II* for German. The occurrences of these tenses in the corpora will be analyzed from the standpoint of an enunciative approach that focuses on the circumstances of the speaker's utterance. For that purpose, we will apply the triadic model of analysis proposed by Rossari (2016) that provides a new conceptualization of how forms conveying modal meaning operate.

This chapter is organized as follows: we first present the triadic model of analysis, giving some illustrations of how the future tense and its temporal and modal meanings are approached in this framework (Section 2), then we present the comparable corpora and the quantitative data we obtained regarding the nuances conveyed by French and German future tenses (Section 3), before summarizing the direct results of this analysis and discussing the degree to which future tense use in newspapers can be said to be modal (Section 4).

2. Qualitative analysis

Our qualitative analysis is based on Rossari, Ricci and Siminiciuc (2017) who identify eleven nuances that may or may not be conveyed by future tenses in French, Italian and Romanian.² Their main observation is that, despite its standard temporal use, future tense differs importantly in the nuances of its modal use in those three Romance languages. Likewise, the uses of the German future tense can be separated into the category of temporal use vs. the category of modal use, but the nuances of meaning do not occur in the same way as in French. The way in which these uses and their semantic nuances vary cross-linguistically will be explained in accordance with the triadic model of analysis (presented in Rossari 2016: 130–131 and Ricci et al. 2016: 96–98) that centers around the speaker's attitude towards the content uttered and the circumstances of the utterance.

2.1 Representation of future tense uses

The representation of the future tense – as a form that can have modal meaning – within the triadic model of analysis differs from traditional descriptions of modality in so far as modal forms are considered not to be intrinsically polysemous and

2. In Rossari et al. (2017), the following eleven nuances are identified and attributed to the future tense in French (FR), Italian (IT) and/or Romanian (RO): epistemic nuance (FR, IT, RO); concessive nuance (IT, RO); emphatic nuance (IT, RO); enunciation nuance (FR, IT, RO); discovery nuance (FR); historical nuance (FR, IT); recapitulative nuance (FR, IT); jussive nuance (FR); mitigative nuance (FR); preterition nuance (FR, IT); quotation nuance (RO). Cf. Azzopardi & Bres (2015) for a detailed analysis of enunciation and discovery nuances.

modal meaning to be semi-conventional. The three components taken into account by this model are the propositional content of an utterance, the enunciation (i.e. the occurrence of an utterance), and the circumstances in which an utterance occurs (i.e. the circumstances of utterance).

Modal forms are described as fundamentally non-polysemous since it is possible to define a semantic core for any linguistic form. The semantics of modal forms are not understood as being monosemous, though: their meaning is unique on an abstract semantic level only, whereas in context, several semantic nuances can derive from the core meaning depending on the circumstances of utterance. The abstract semantic indication is stable for all uses of a given form, regardless of the different nuances, but it can operate on two different levels, either on the propositional content or on the enunciation. Not all linguistic forms are provided with the potential to operate in both ways (see following paragraph). It is precisely the possibility to function on the enunciative level that – combined with the specific circumstances of utterance, including context – enables the emergence of modal nuances. Thus, the nuances of the future tense are not understood as resulting from an *a priori* temporal or modal meaning with a direct involvement of contextual data to this meaning, but as primarily depending on the level on which the common core meaning operates. To be more precise, a semantic indication of posteriority underlies all future tense uses, but it applies to the propositional content for future tense forms used temporally (1st component), whereas it applies to the enunciation for modal future tense forms (2nd component). Both when this indication operates on the propositional level and on the enunciative level, the specific circumstances in which a given utterance occurs specify different temporal and modal nuances of the future tense (3rd component). When the realization of the state of affairs expressed in the utterance is considered to occur in the future, the speaker represents the event denoted by the predicate, by means of the future tense, as being posterior to the time of utterance. The future tense is then used as a temporal marker fulfilling particular communicative functions, and conveys different nuances that vary with the circumstances of utterance. But when the realization of the state of affairs is excluded from occurring in the future because the predicate refers to a past or present event, the use of future tense in such circumstances does not allow its actualization as a tense, but as an enunciative marker – fulfilling particular rhetorical functions. The indication of posteriority then applies to the enunciation, with the speaker evoking an utterance, which would fictionally occur after the time of utterance. This process can be related to a rhetorical strategy, similar to preterition: The speaker utters a given content while representing, by means of the future tense, the upcoming occurrence of another utterance having the same content. This strategy helps the speaker to avoid endorsing the content at the time of utterance, motivated by the fact that his or her utterance is a matter of conjecture, assessment, and so on.

In the same way as for temporal future tense, the different modal nuances depend on the circumstances in which the future tense is used.

The semi-conventional feature of modal meaning is linked to precisely those circumstances. Whereas the ability of a given form to operate on the enunciative level is conventional – although not all forms have this potential – the circumstances that specify nuances of meaning are not. In fact, one needs to contextualize the given form. For example, temporal meaning is excluded for the future tense in the utterance *Elle aura bien célébré ses 30 ans hier* ('She (will have) celebrated her 30th birthday yesterday'), obviously referring to a state of affairs in the past and thus having modal meaning, namely an epistemic nuance or a recapitulative nuance. Both nuances correspond to a retrospective judgment on the event mentioned in the utterance; the former shows a judgment on the actual realization of the event, based on inference ('probably celebrated'), the latter shows a judgment on the event itself, based on awareness of the facts ('finally celebrated'). But without any further information on the circumstances of utterance – indicating whether the speaker's utterance is a matter of conjecture or assessment of a past event according to his or her knowledge about the content uttered – the actual modal nuance conveyed by the French future tense cannot be determined. Yet, an equivalent utterance with the future tense in German (*Sie wird gestern ihren 30. Geburtstag gefeiert haben*) can only express a conjecture, since the future tense is not used for retrospective assessments. Temporal nuances of the future tense are also determined by the circumstances of utterance: Whereas the French utterance *L'examen se présentera sous forme de QCM* ('Examination will/shall be a multiple-choice test') features either a predictive or a directive content, the equivalent utterance in German (*Die Prüfung wird in Form eines Multiple-Choice-Tests sein*) can only be interpreted as a prediction, since the future tense is excluded from directive contents, such as in official guidelines (Fritz 2000: 56).

2.2 Future tense in French and German

Based on the observations of Rossari et al. (2017) on nuances of the future tense for three Romance languages (see footnote 2), we will give a brief description of those nuances that are relevant for the French and German future tenses, following the triadic model of analysis introduced above. Authentic examples from our French corpus will help us for this description; examples in German will be used only to illustrate significant differences between the two languages. The paradigm under consideration is the French *futur simple* and *futur antérieur* vs. the German *Futur I* and *Futur II*.

2.2.1 Temporal use of future tense

In temporal uses of the future tense, the indication of posteriority operates on the predicate of the proposition.³ We distinguish three different nuances of the temporal future tense, depending on the circumstances of utterance (see Figure 1).

In its standard use, the future tense conveys a prospective nuance, meaning the speaker predicts an as yet unrealized state of affairs to occur after the time of utterance (Bybee, Perkins & Pagliuca 1994: 244; Duden 2009: § 729,732; Grevisse & Goosse 2016: § 887–888). This can be the case, for example, for a scheduled event (1), an intended goal (2), an estimated evolution (3), or an expected reaction (4), sharing some characteristics, including the speaker's lack of knowledge about these events at the time of utterance:

- (1) *L'événement aura lieu le 25 août à Essen.* (LM: 22/06/07)⁴
'The event is going to take place in Essen on the 25th of August'.
- (2) *J'ai été élu, je réussirai.* (LM: 25/10/07)
'I was elected, I will succeed'.
- (3) *Dans le sud de l'Europe, la disponibilité en eau baissera de 5% à 35% [...].*
(LM: 19/03/07)
'In Southern Europe, water availability will decrease by 5% to 35%[...]'
- (4) *La cuisine arménienne ravira les amateurs de saveurs orientales.*
(LM: 02/02/07)
'Armenian cooking will delight lovers of oriental flavors.'

The future tense with a prospective nuance actually covers a wide range of contexts related to prediction. This variation is partly due to the degree of possibility that the speaker associates with the state of affairs to happen. Concerning this possibility feature, one actually observes that the most salient adverbial collocates of the future tense by log-likelihood in a right-context window of five words are not only temporal adverbs like *bientôt* 'soon', *demain* 'tomorrow', *(plus) tard* 'later' and *prochainement* 'shortly', but also epistemic adverbs like *peut-être* 'possibly', *probablement* 'probably', *certainement* 'most probably' and *sûrement* 'very probably' (see Figure A, Appendix). Whereas the adverb *peut-être* can be added to (4), the other examples seem to accept more easily adverbs expressing a higher degree of possibility like *probablement* (3) or even *sans aucun doute* 'definitely' (1–2), depending on the speaker's attitude towards the content uttered. Thus, occurrences of the future

3. Conditional uses of the future tense (i.e. in constructions such as *si P, Q* 'if P, Q') are counted as part of the temporal category.

4. Corpus references indicate the newspaper and its publication date, for example LM: 22/06/07 meaning 'Le Monde issued 22 June 2007' or SZ: 22/06/13 meaning 'Süddeutsche Zeitung issued 22 June 2013'.

tense with a prospective nuance concern states of affairs that are considered to form a continuum ranging from the most to the least likely to happen. To model the use of the future tense with a prospective nuance, we propose the following scheme:

CIRCUMSTANCES_(IGNORANCE): SPEAKER – UTTER – [SUBJECT – PREDICATE] →
PROSPECTIVE NUANCE

which means that the speaker refers to an as yet unrealized state of affairs that he or she does not know about, while assuming its likelihood to occur after the time of utterance. The German future tense behaves similarly.

In addition to the prospective nuance, the future tense in its temporal use may also, albeit less often, feature a *jussive* nuance. This nuance is conveyed, for example, when expressing a command (5), a rule (6), a recommendation (7), or an invitation (8).

- (5) *Tu aimeras ton prochain comme toi-même.* (LM: 10/11/07)
‘You shall love your neighbor as yourself’.
- (6) *Pourront toutefois être inscrites auxdits budgets les dépenses relatives à des services d’aumônerie [...].* (LM: 17/07/07)
‘Expenditures on chaplaincy services, however, may be entered into the aforementioned budget [...]’.
- (7) *[...] les tiges trop longues seront raccourcies de façon qu’elles ne dépassent plus.* (LM: 11/10/07)
‘[...] stems that are too long shall be shortened so that they won’t stick out’.
- (8) *On lira notamment la contribution de Gwyn Richard Campbell [...].* (LM: 17/05/07)
‘One should refer in particular to Gwyn Richard Campbell’s contribution [...]’.

These examples express order *sensu lato*, with the speaker calling for compliance in the future. They differ from the standard use of the future tense, although their use remains temporal, because the speaker’s intention is not to indicate the likelihood that a state of affairs will occur after the time of utterance. We rather understand these examples as the speaker representing a state of affairs that he or she requires to occur. Such a required event will necessarily be pictured as occurring after the time of utterance and therefore be temporally bounded to the future. Thus, we propose the following scheme to represent the jussive nuance of the future tense:

CIRCUMSTANCES_(REQUIREMENT): SPEAKER – UTTER – [SUBJECT – PREDICATE] →
JUSSIVE NUANCE

which means that the speaker refers to an as yet unrealized state of affairs that he or she requires to occur after the time of utterance. Unlike the above examples of

the prospective future tense (1–4), we did not observe specific lexical collocates for the jussive future tense, but specific grammatical features like second-person or generic subjects (*tu/vous* ‘you’, *on* ‘one’, *la/le(s) X* ‘the X, referring to an entire group of things or people’) and passive constructions. This might suggest that the necessity feature of the jussive future tense, in the form of grammatical items, is an integral part of the propositional content in contrast to the prospective future tense where the possibility feature expressed by epistemic adverbs is optional and extra-propositional. The German future tense partly differs from this description of the jussive nuance since this particular use seems to be much more restricted. Neither legal nor instructional texts (laws or contracts; user guides or rules of play) resort to this type of future tense, which seems to be limited to direct orders during verbal exchange, like *Du wirst sofort dein Zimmer aufräumen* ‘You’re going to tidy up your room right now’; invitations, such as French *Tu salueras bien ta mère* ‘Please give my regards to your mother’, are excluded as well. Given these restrictions and the text type analyzed, it is hardly surprising that we did not find any occurrences of this future tense use in the German corpus.

Following Rossari et al. (2017), we propose a third nuance of the temporal future tense, the mitigative nuance, which is as rare as the jussive nuance. This particular use, shared with the prospective and jussive nuances, provides an indication of posteriority that operates on the predicate. The following examples show how it differs:

- (9) *Je ne parlerai pas de “maillon faible”, cela serait insultant.* (LM: 22/02/07)
‘I would not speak of a “weak link”; it would be insulting’.
- (10) *On a assisté à un moment tout à fait important, j’oserai dire historique.* (LM: 31/08/07)
‘We attended a very important, I would dare say historical, moment’.
- (11) *Je répondrai en affirmant, d’abord, le droit des musulmans, des bouddhistes ou des juifs à célébrer librement leurs fêtes.* (LM: 22/12/07)
‘I would answer by affirming, first of all, the right of Muslims, Buddhists or Jews to celebrate their holy days freely’.
- (12) *Pour camper le décor, on rappellera que la physique vit, depuis bientôt un siècle, une forme de schizophrénie.* (LM: 07/06/07)
‘To set the scene, one should recall that physics has been experiencing, for almost a century, a sort of schizophrenia’.

Regarding their propositional content, these examples might seem to be too heterogeneous to be merged into a single category. However, functionally, the use of the future tense can be linked to one and the same purpose (similar to rhetorical preterition): artfully drawing attention to a matter by pretending not to mention

it. Formally, all occurrences of this future tense type in the French corpus have a first-person singular or a generic subject (*je* 'I', *on* 'one'). The predicate (that can be simple or complex, e.g. *oser dire* in (10)) is a *verbum dicendi* and can be schematized as 'say/demand/answer (that) X'. The complement of this predicate expresses a potentially offensive content – in the sense that it could be face-threatening (9) or not commonly known or accepted (10–12). The speaker represents his or her act of saying this offensive content, by means of the future tense, as forthcoming, in order to distance him- or herself from it.⁵ We found the following examples in the German corpus:

- (13) *Man wird nicht in Abrede stellen können, dass Hanna Rosins Beobachtungen jeweils stimmen.* (SZ: 26/01/13)
 'One could not deny that each of Hanna Rosin's observations are right'.
- (14) *Aber man wird Geißendörfer [...] nicht vorwerfen können, mit seiner Serie nichts bewegen zu wollen.* (SZ: 11/10/13)
 'But one could not accuse Geißendörfer [...] of not trying to change things with his series'.
- (15) *Na ja, träumen wird man ja noch dürfen.* (SZ: 30/01/13)
 'Well, one should be free to dream'.

As in the French examples, the speaker uses the future tense to represent an as yet unrealized state of affairs. In this way, he or she wants to tone down a potentially polemical value judgment (13–15). The use of the generic subject *man* 'one', negation or modal particles like *ja* (signifying that the content should already be known) in addition to the future tense helps to qualify a statement that would be too assertive with the systematic presence of a root-modal verb. In consideration of these functional similarities in French and German, we opt for one single category, outlined in the following scheme:

CIRCUMSTANCES_(OFFENSE): SPEAKER – UTTER – [SUBJECT – PREDICATE] →
 MITIGATIVE NUANCE

which means that the speaker refers to a state of affairs that is potentially offensive to mention, while representing his or her act of saying to occur after the time of utterance to be less assertive.

5. We would have included utterances like *Cela vous fera 25€, s'il vous plaît* ('That will be 25€, please') in the category of mitigative future tense as well, but there were no such occurrences in the corpora. Even though differing formally from the above examples (9)–(12), the function is the same, since the content uttered is represented to occur in the future in order to avoid a face-threatening act.

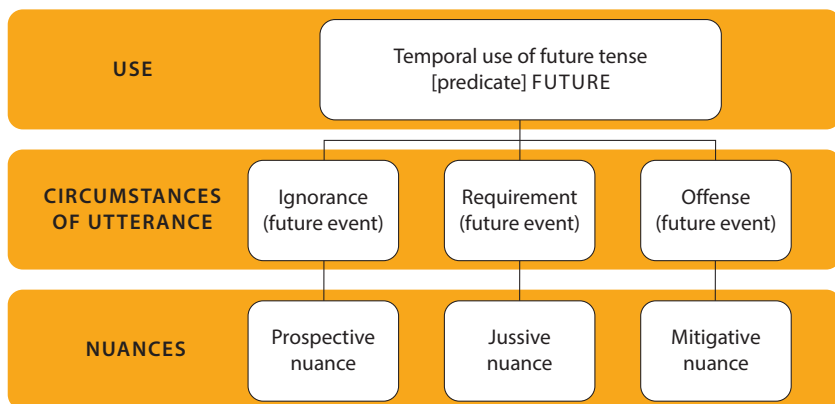


Figure 1. Temporal use of the future tense and the nuances it conveys in specific circumstances of utterance

2.2.2 Modal use of future tense

We distinguish the modal use of the future tense from its temporal use. They differ in that only the former functions enunciatively. Whereas temporal nuances of the future tense occur in specific circumstances and serve particular communicative functions (i.e. predicting, commanding, toning down), modal nuances differ in the circumstances of utterance and serve rhetorical functions (see Figure 2). Two possible circumstances may trigger the enunciative function of the future tense: either the speaker is ignorant of the realization of the state of affairs mentioned in his or her utterance and referring to the past or present (epistemic nuance), or he or she knows about the outcome of the state of affairs (historical and recapitulative nuances). In both cases, the future tense becomes an enunciative marker, with the indication of posteriority applying to the enunciation and not to the propositional content – meaning that it evokes an upcoming utterance having the same content as the one uttered – but the underlying motivation for this rhetorical strategy is not the same.

The epistemic nuance of the future tense occurs when the speaker talks about a past or present state of affairs without knowing whether it actually occurs or has occurred. The strategy of showing the enunciation as fictionally forthcoming, by means of the future tense, enables the speaker to avoid endorsing an uncertain content at the time of utterance.

- (16) *Ce détail n'aura évidemment échappé à personne.* (LM: 26/11/07)
 'This detail, of course, will not have escaped anyone's attention.'
- (17) *Vous aurez noté ma bienveillance à ne pas souligner vos propos malheureux sur cette banlieue où je vis [...].* (LM: 18/01/07)
 'You will have noticed that I have been good enough not to highlight your unfortunate remarks on the suburbs where I'm currently living [...].'

- (18) *Selon les résultats définitifs publiés, lundi 23 avril, par le ministère de l'intérieur, François Bayrou, avec 6 820 914 voix, obtient 18,57% des suffrages. [...] Il ne regrettera pas son court déplacement à Mayotte puisqu'un quart (25,67%) des voix des Mahorais se sont portées sur lui.* (LM: 24/04/07)
 'According to the final results released by the Interior Ministry on Monday, April 23rd, François Bayrou, with 6,820,914 votes, receives 18.57% of the votes. [...] He will not regret his short trip to Mayotte since one quarter (25.67%) of Mahorans voted for him'.
- (19) *Les données du problème sont donc limpides [...]. Compte tenu de leur population – la Chine, l'Inde et l'Indonésie comptabiliseront 2,5 milliards d'individus – [...], les pays émergents font que la planète court à la catastrophe.* (LM: 12/02/07)
 'The factors in the problem are crystal clear [...]. Given their population – China, India and Indonesia will reach 2.5 billion individuals – [...], the planet is heading for disaster because of the emerging countries'.

All these examples are conjectures about a state of affairs, referring to the past (16–17) or to the present (18–19). Example (18) in particular illustrates that this type of future tense is used when the speaker's assumption relies on evidence and is based on inference (*puisque* 'since').⁶ None of these examples is compatible with temporal indications like *demain* 'tomorrow' or *à l'avenir* 'in the future' without changing the meaning of the utterance. To represent this modal use, which is the same in German, we propose the following scheme:

CIRCUMSTANCES_(INFERENCE): SPEAKER – UTTER – [SUBJECT – PREDICATE] →
 EPISTEMIC NUANCE

which means that the speaker refers to an ongoing or past state of affairs that he or she infers from evidence, while explicitly avoiding endorsing the content of the utterance, which is a matter of conjecture.

The historical and recapitulative nuances both appear in those circumstances in which the speaker refers to a state of affairs – either in reality or in fiction – whose outcome he or she is aware of. The contexts can be various from featuring life stories (20), fictional stories (21), reports (22) or commentaries (23):

- (20) *Durant la seconde guerre mondiale, il effectuera son service militaire [...].*
 . (LM: 30/10/07)
 'During World War II, he performed his military service [...]'

6. Cf. Azzopardi (2011: 401–402, 413–419) for a definition and a detailed state of the art of the so-called *futur de conjecture* ('conjectural future') and the underlying inferential process.

- (21) *Il n'ira pas à son bureau ce matin, ni le lendemain ni aucun autre matin.*
(LM: 30/08/07)
'He was not to his office this morning, nor the next day nor any other morning.'
- (22) *Elle [la maladie] aura, au total, atteint plus de 2 000 élevages de bovins et d'ovins.*
(LM: 09/01/07)
'In total, the disease affected more than 2,000 cattle and sheep farms.'
- (23) *Jacques Chirac aura donc quitté le pouvoir comme il l'a exercé pendant douze ans [...].*
(LM: 16/05/07)
'Jacques Chirac left office the same way he exercised power for twelve years [...].'

With none of these future tense forms does the speaker refer to events posterior to the time of utterance, but the future tense is rather used for the reconstruction of events. Yet, these examples differ with regard to the speaker's perspective: The speaker focuses on a past event either from event time – which explains anaphoric elements like *lendemain* 'next day' – (20–21, historical nuance) or from the time of utterance (22–23, recapitulative nuance). This difference explains why the underlying rhetorical strategy expressed by the future tense is not the same. Within narrative discourse, the speaker stages him- or herself as a witness of the time he or she relates, and the strategy of showing the enunciation as forthcoming serves to reconstruct the chronological succession of events with the benefit of hindsight. Within non-narrative discourse, the fictive transfer of enunciation enables the speaker to leave some doubt as to his or her awareness of the content: The utterance being a matter of retrospective assessment, the speaker uses the future tense to make it less categorical, because the content is based either on second-hand information (22) or on opinion and not facts, that is, value judgment (23). We present the two nuances as follows:

CIRCUMSTANCES_(AWARENESS): SPEAKER – UTTER – [SUBJECT – PREDICATE] →
HISTORICAL NUANCE / RECAPITULATIVE NUANCE

which means that the speaker refers to a past state of affairs that he or she is aware of, while avoiding explicitly endorsing the content of the utterance, which is a matter of narration or assessment. German functions similarly, in narrative cases, but differently, in non-narrative cases. Actually, the future tense in German can convey a historical nuance within narrative discourse, but there is no recapitulative nuance, because the rhetorical strategy of being less assumptive for an utterance about past events would automatically engender an epistemic reading.

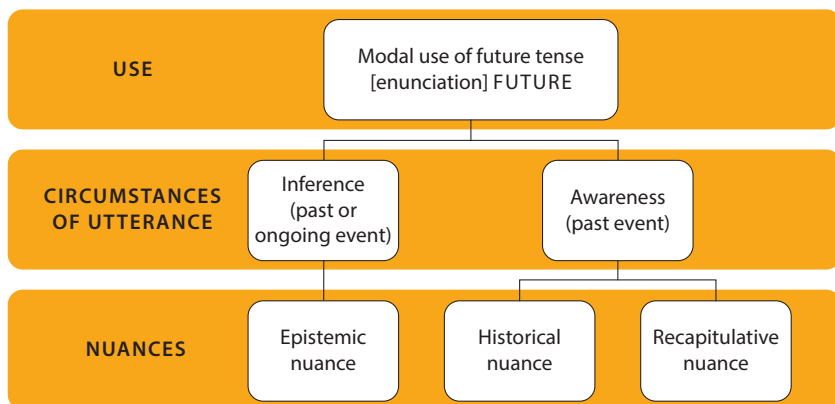


Figure 2. Modal use of the future tense and the nuances it conveys in specific circumstances of utterance

According to our qualitative analysis, French and German future tenses do not differ significantly in conveying temporal and modal nuances. To put this preliminary observation in perspective, we will adopt a quantitative approach in order to elicit statistical evidence on the modal and temporal use of the future tense in French and German daily newspapers.

3. Quantitative analysis

Our quantitative analysis is based on data from comparable corpora of French and German newspaper texts. This particular genre seems to lend itself well to research on modality for two reasons. First, it encompasses various subgenres offering both factual information (i.e. news reports) and personal opinion (i.e. editorials, reviews). Second, it merges different modes of discourse (i.e. descriptive, narrative or argumentative). To complement the previous qualitative observations on the uses of the future tense, we propose a systematic corpus study that provides statistical evidence on the temporal and modal nature of the future tense in the analyzed genre. We will give the frequency of occurrence and the proportion of each temporal and modal nuance based on the above classification. We need to point out that our quantitative data reflect the features of the future tense such as it appears in the corpora, without any extrapolation of this information to language in general. The underlying interest is to highlight tendencies concerning modality in a particular journalistic genre that go beyond qualitative assumptions.

3.1 Corpora and procedures

For each language, the working corpus contains an annual collection of newspaper issues. The corpus consists of two daily newspapers that are similar not only in terms of circulation and thematic structure, but also in terms of dates and size: *Le Monde* from 2007 for French (= *LM07*: 22,126,082 tokens) and *Süddeutsche Zeitung* from 2013 for German (= *SZ13*: 21,996,944 tokens).⁷ We accessed these corpora via two different platforms: *BTLC* for the French corpus and *COSMAS II* for the German corpus.⁸ Both platforms allow the user to generate concordances based on specific queries, to compile statistics on lexical or grammatical occurrences and co-occurrences, as well as to create subcorpora related to particular text sections or themes.

While the query for future tense forms is unproblematic in the French corpus – each verb has a tag indicating whenever it is conjugated in the future tense – one has to type a complex query in the German corpus because of the periphrastic form of the German future tense, which is not tagged.⁹ Both corpora have too many future tense occurrences to be examined manually; we therefore opted for three subcorpora that could be analyzed exhaustively: the first subcorpus representing a random sample with 5% of all occurrences from the entire corpus, a second subcorpus of all occurrences in the editorial section (*LM07*: 643,458 tokens; *SZ13*: 501,539 tokens) and a third subcorpus of all occurrences in the science section (*LM07*: 515,966 tokens; *SZ13*: 542,369 tokens). These subcorpora serve a double purpose: first, to verify that the quantitative data observed in the random sample

7. Given the tagged corpora that are available on the platforms used, we could not choose identical dates for both newspapers.

8. *BTLC* <persan.rom.uni-koeln.de/btlsc> is a restricted-access platform containing corpora from a Franco-German research project on the diachronic evolution of French prepositions (*PRESTO*, cf. <prestos.ens-lyon.fr>); *COSMAS II* <cosmas2.ids-mannheim.de/cosmas2-web> is a free-access platform provided by the *Institut für deutsche Sprache* (Mannheim, Germany) containing German reference corpora (cf. <ids-mannheim.de/kl/projekte/korpora>).

9. However, this complex query in the German corpus, which is not tagged for future tense, generates noisy concordance results, with about one third of false positives, which were removed manually in each subcorpus. Characteristic noise concerns syntactical disjunctions of the auxiliary *werden* and the infinitive because of the chosen span of 20 items, or passive constructions because of the homograph *werden* and past participles wrongly tagged as infinitives. There are also tagging errors where (proper) nouns or other categories ending with *-en* are considered to be infinitives. Even though noise related to tagging errors exists in the French subcorpora as well, it represents less than 1% for each. For the time being, the French periphrastic future tense – *futur proche* – has not been integrated to the paradigm under consideration because it is not tagged either and thus would imply similar time-intensive work to that involved with the German future tense.

roughly coincides with those from the other subcorpora in order to avoid, as far as possible, corpus biases; and second, to highlight significant differences in frequency related to subgenres or modes of discourse.

After corpus queries and concordance extraction, all occurrences were labeled manually by identifying the actual nuance being conveyed, based on contextual data, in accordance with the above classification.¹⁰ The resulting table of absolute frequencies was, on the one hand, totaled for each subcorpus so that the quantity of each use or nuance in every subcorpus could be displayed (see Figures 3, 6, 7) and, on the other hand, transformed into relative frequencies so that the number of occurrences of each use or nuance in the different subcorpora could be compared (see Figures 4, 5).

3.2 Quantitative data

The quantitative analysis reveals some interesting information on modality and the future tense in the corpora, starting with the fact that modal use of the future tense appears to be rather rare in both languages compared to their temporal use as displayed in Figure 3.

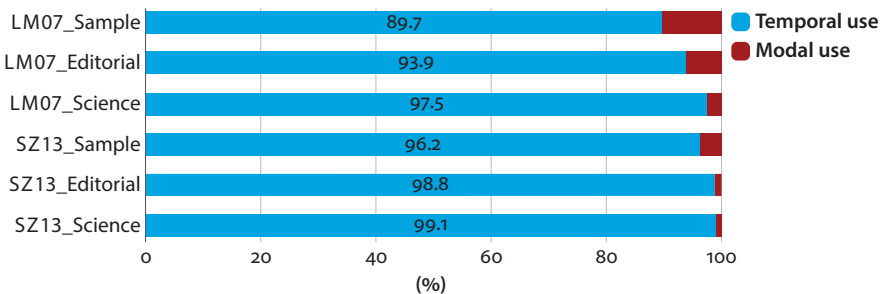


Figure 3. Percentage of temporal and modal uses of the future tense (French corpora: *LM07_Sample*, *LM07_Editorial*, *LM07_Science*; German corpora: *SZ13_Sample*, *SZ13_Editorial*, *SZ13_Science*)

Figure 3 indicates that the amount of temporal use of the future tense varies from 90% to 98% in the French subcorpora, and even more, from 96% to 99% in the German subcorpora, reducing the modal future tense to a very small proportion in both languages. In addition, one observes a similar decrease of modal use for the three pairs of subcorpora: the respective science sections display a lower percentage

10. Cf. Heine (1995: 25–36) for a discussion of some contextual factors being relevant, besides linguistic elements, for characterizing the uses of modal forms.

of such uses than the editorial sections and the random samples. The latter, as representative excerpts of the entire corpora, are expected to reveal the average use of the modal future tense in the respective newspapers. Both random samples have a higher percentage of the modal future tense than the chosen thematic sections (editorial and science), which would mean that modal uses of the future tense are under-represented in these sections compared to the entire corpus. Overall, German has a systematically lower use of the modal future tense than French in the newspaper texts analyzed. This tendency is confirmed when we compare the relative frequencies generated for both uses as displayed in Figure 4.

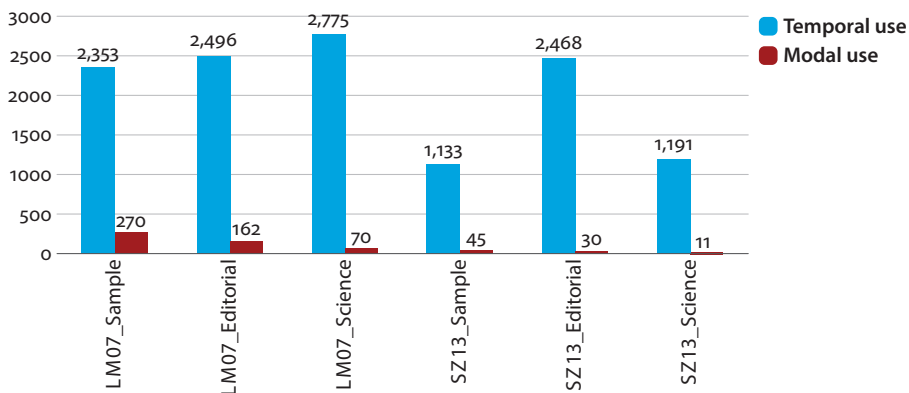


Figure 4. Relative frequencies (per million words) of temporal and modal uses of the future tense (French corpora: *LM07_Sample*, *LM07_Editorial*, *LM07_Science*; German corpora: *SZ13_Sample*, *SZ13_Editorial*, *SZ13_Science*)

Indeed, Figure 4 confirms that there is a significant difference in the number of the temporal and of the modal future tense occurrences in each subcorpus, modal use being rather infrequent in both languages. For example, the French random sample reveals that the temporal future tense appears 2,353 times per million words in the newspaper *Le Monde* from 2007, whereas the modal future tense only appears 270 times per million words. Yet, the German random sample only shows 1,133 temporal and 45 modal occurrences of the future tense per million words. The tendency of German subcorpora to have lower frequencies of both future tense uses compared to the French subcorpora is almost constant, the only exception being the German editorial subcorpus. The significantly higher frequency of the temporal future tense in this subcorpus compared to its frequency in the random sample would suggest that editorials form a particular subgenre within newspapers, whereas the science section does not, duly corresponding to the frequencies of the random sample. For all French subcorpora, the relative frequencies of the temporal

future tense are roughly similar as well. Finally, the variation of the modal future tense occurrences between the three subcorpora in each language, on the one hand, and between French and German corpora, on the other hand, can be represented in terms of the frequency of each modal nuance as in Figure 5.

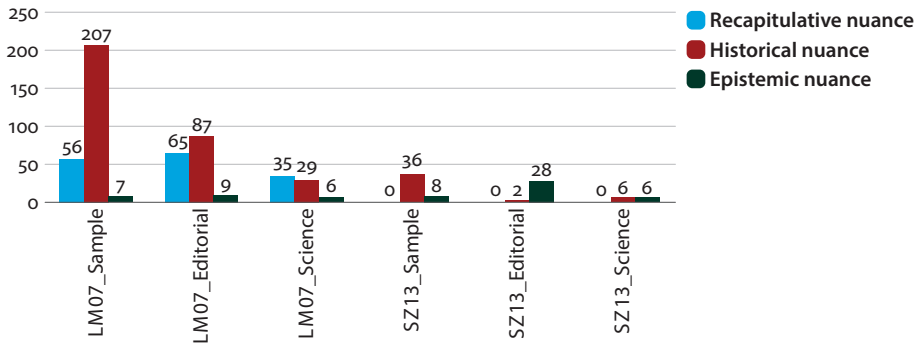


Figure 5. Relative frequencies (per million words) of recapitulative, historical and epistemic nuances of the modal future tense (French corpora: *LM07_Sample*, *LM07_Editorial*, *LM07_Science*; German corpora: *SZ13_Sample*, *SZ13_Editorial*, *SZ13_Science*)

Figure 5 is a close-up of the total of modal occurrences given in Figure 4.¹¹ It shows that the variation of the modal future tense occurrences, which appears both on inter-lingual and intra-lingual levels, is related to variations of modal nuances depending on specific contexts. Concerning the intra-lingual level, variations in French do not concern epistemic or recapitulative nuances, whose frequencies are very similar, with the random sample being representative of the two other subcorpora, but it concerns the historical nuance. Actually, historical future tense occurrences are significantly under-represented in the editorial and science sections compared to the entire corpora. This seems to be mostly due to the fact that these sections do not feature narrative discourse as much as the random sample does (especially in book or movie reviews from the culture section). Variations in German seem to result from the discursive properties of each subcorpus as well. The lack of modal future tense occurrences in the German corpora can be related to the principal informative function of newspaper articles, especially obvious in the science section. However, the expression of personal opinion increases the frequency of epistemic occurrences in the editorial section. Narration using historical future tense is featured in different sections represented in the random sample. On

11. Any differences between the individual modal occurrences in Figure 5 taken together and the total of modal occurrences given in Figure 4 results from the figures being rounded to the nearest ten.

the inter-lingual level, these variations of modal nuances are due to the fact that the recapitulative nuance does not exist for the German future tense and that historical future tense does not seem to be a common tense in narrative discourse in German. Given those variations, the German subcorpora appear to be less modal than the French ones concerning the use of the future tense.

Finally, we will consider the proportion of each nuance within the temporal and the modal future tenses displayed in Figures 6 and 7. Both are close-ups of the distributional data given in Figure 3. Figure 6 shows that the temporal future tense is dominated by the amount of prospective nuance in each subcorpus, reaching even 100% in the German editorial and science subcorpora. The jussive nuance is exclusive to the French corpora, even though it never exceeds 1%. The mitigative nuance is used in similarly small quantities in the French subcorpora and, for German, only appears in the random sample. These tendencies, however, might be specific to the text type. Thus, the temporal future tense is almost exclusively used

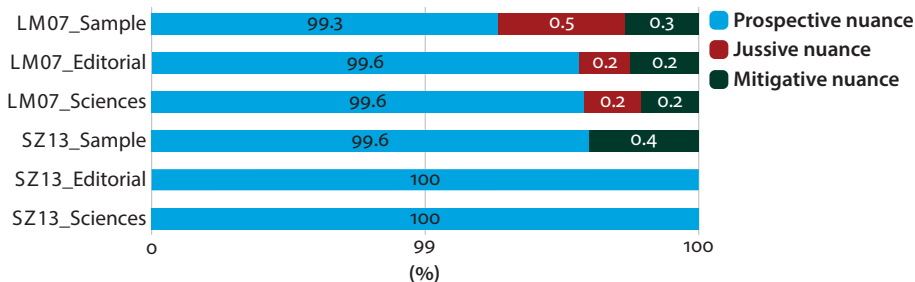


Figure 6. Percentage of prospective, jussive and mitigative nuances of the temporal future tense (French corpora: *LM07_Sample*, *LM07_Editorial*, *LM07_Science*; German corpora: *SZ13_Sample*, *SZ13_Editorial*, *SZ13_Science*)

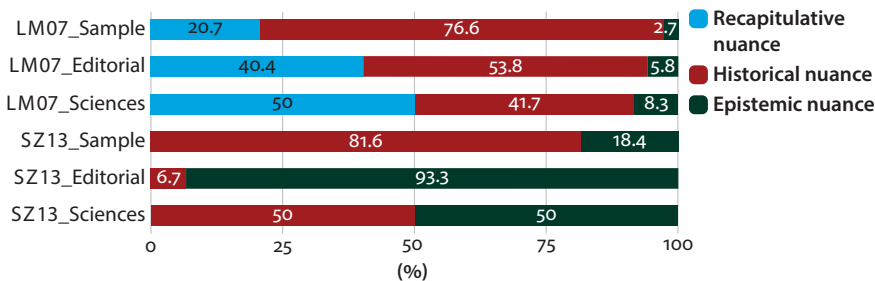


Figure 7. Percentage of recapitulative, historical and epistemic nuances of the modal future tense (French corpora: *LM07_Sample*, *LM07_Editorial*, *LM07_Science*; German corpora: *SZ13_Sample*, *SZ13_Editorial*, *SZ13_Science*)

to make predictions relating to future states or events. In contrast, there is considerable variation in the volumes of modal nuances, as shown in Figure 7, so that it is not easy to notice inter-lingual or intra-lingual tendencies. The only conclusion that may be drawn from Figure 7 is that there is a three-way distribution of modal nuances for the French subcorpora but only a binary distribution for German subcorpora as German lacks the recapitulative future tense.

4. Discussion

We will recapitulate the direct results of our analysis by discussing the degree of modality of the future tense in newspapers. Indeed, combining quantitative with qualitative analysis has made possible to learn more about the modal future tense in both French and German as well as about the variations of modal nuances that appear cross-linguistically.

Let us first summarize the results of the quantitative analysis of the future tense: Corpus data show that (i) the rate of modal use is very low in both languages compared to the temporal use of the future tense, (ii) the frequency of future tense occurrences, for both modal and temporal use, is generally lower in German than in French, (iii) the frequency of the individual modal nuances conveyed by the future tense is widely disparate, within one language as well as inter-lingually, and (iv) the variation in the number of nuances conveyed by the future tense is significant only for modal use but not for temporal use.

In addition, the corpus analysis shows some interesting tendencies among newspapers and their thematic sections with regard to modality and future tense. Given the primarily informative function of newspapers, (i) the future tense is almost exclusively used to make predictions about future events and so modal use is rare, (ii) historical future tense is restricted to specific discursive contexts (i.e. book or film reviews), that would be particularly incompatible with the science section, and (iii) epistemic future tense is restricted to specific discursive contexts (i.e. expression of opinion), that seem to be particularly compatible with the editorial section.

All in all, our corpus analysis has allowed us to go beyond the previous qualitative hypothesis on the use of the future tense, that is, of certain modal nuances existing or missing in the analyzed languages. In fact, the integration of quantitative methods informed by our preliminary observation, according to which the French and German future tenses do not differ significantly in conveying temporal and modal nuances, enables us to draw the following expanded conclusion: The modal use of the future tense is a rare phenomenon in both French and German, at least

in the text type sampled. But on the basis of our comparison – given the greater amount of modal rather than temporal use, higher relative frequencies of modal occurrences and a greater range of modal nuances – the modal use of the future tense appears more frequent in French than in German.

5. Conclusion

This chapter investigated modal meaning of future tenses in French and German based on occurrences in daily newspapers. Corpus data revealed interesting aspects in addition to the preliminary qualitative analysis, according to which modal use of the future tense derives from an enunciative function of these forms and the future tense differs in the nuances of meaning conveyed in French and German (i.e. jussive and recapitulative nuances). Modal use of the French *futur simple* and *futur antérieur* or of the German *Futur I* and *Futur II* turned out to be relatively rare compared to their temporal use, so that these grammatical items remain, above all, temporal and not enunciative markers. According to our quantitative analysis, French appears to have a more modal use of the future tense than German. In order to investigate whether this tendency is confirmed with respect to lexical items, it would be interesting to complete our contrastive research by analyzing the nature and degree of modality expressed by French and German modal verbs and the auxiliary verb *aller* ‘to go’, used in the French periphrastic future tense (*futur proche*). As to the paradigm of modal verbs, we would expect to find a more frequent use and a greater range of nuances they convey in the German corpus compared to the French one.

Acknowledgements

This study is part of the research project *La représentation du sens modal et de ses tendances évolutives dans deux langues romanes: le français et l’italien* (‘Representing modal meaning and its evolutionary trends in two Romance languages: French and Italian’), directed by Corinne Rossari and funded by the Swiss National Science Foundation, project no. 100012_159458, cf. <<http://p3.snf.ch/Project-159458>>.

References

- Azzopardi, Sophie. 2011. *Le futur et le conditionnel: Valeur en langue et effets de sens en discours. Analyse contrastive Espagnol/Français*. PhD dissertation, University of Montpellier 3.
- Azzopardi, Sophie & Bres, Jacques. 2015. Quand le futur ne porte pas sur le procès qu'il actualise: Futur d'énonciation et futur de découverte. *Revue de Sémantique et Pragmatique* 38: 77–96.
- Bybee, Joan L., Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World*. Chicago IL: The University of Chicago Press.
- Celle, Agnès. 2006. *Temps et modalité. L'anglais, le français et l'allemand en contraste*. [Études Contrastives 7]. Bern: Peter Lang.
- Confais, Jean-Paul. 1992. No future? Les 'futurs' du français et de l'allemand. In *Systèmes interactifs. Mélanges en l'honneur de Jean David*, Gertrud Gréciano & Georges Kleiber (eds), 81–91. Paris: Klincksieck.
- Damourette, Jacques & Pichon, Édouard. 1911–1936. *Des mots à la pensée. Essai de grammaire de la langue française*, Vol. 5. Paris: Éditions d'Artrey.
- Duden = Dudenredaktion (Bibliographisches Institut). 2009. *Duden 4: Die Grammatik*, 8th edn. Mannheim: Dudenverlag.
- Fleischman, Suzanne. 1982. *The Future in Thought and Language. Diachronic Evidence from Romance* [Cambridge Studies in Linguistics 36]. Cambridge: CUP.
- Fritz, Thomas A. 2000. *Wahr-Sagen: Futur, Modalität und Sprecherbezug im Deutschen*. [Beiträge zur germanistischen Sprachwissenschaft 16]. Hamburg: Buske.
- Grevisse, Maurice & Goosse, André. 2016. *Le bon usage: Grammaire française*, 16th edn. Brussels: De Boeck Supérieur.
- Heine, Bernd. 1995. Agent-oriented modality vs. epistemic modality: Some observations on German modals. In *Modality in Grammar and Discourse* [Typological Studies in Language 32], Joan L. Bybee & Suzanne Fleischman (eds), 17–53. Amsterdam: John Benjamins. doi:10.1075/tsl.32.03hei
- Revaz, Françoise. 2009. Valeurs et emplois du futur simple et du présent prospectif en français. *Faits de Langues* 26: 149–161.
- Ricci, Claudia, Rossari, Corinne & Siminiciuc, Elena. 2016. La représentation des sens modaux dans trois langues romanes: Le français, l'italien et le roumain. Du qualitatif au quantitatif et retour. *Syntaxe et Sémantique* 17: 93–113.
- Rossari, Corinne. 2016. Les fluctuations de sens dans quelques formes modales à la lumière d'une approche quantitative et qualitative. *Studii de Lingvistică* 6: 127–142.
- Rossari, Corinne, Ricci, Claudia & Siminiciuc, Elena. 2017. Les valeurs rhétoriques du futur en français, italien et roumain. In *Le futur dans les langues romanes*, Laura Baranzini (ed.), 49–77. Bern: Peter Lang.
- Sundell, Lars-Göran. 2003. Le futur modal revisité. In *Aspects de la modalité*, Merete Birkelund, Gerhard Boysen & Poul Søren Kjærsgaard (eds), 217–228. Tübingen: Niemeyer. doi:10.1515/9783110963434.217
- Vater, Heinz. 1975. Werden als Modalverb. In *Aspekte der Modalität* [Studien zur deutschen Grammatik 1], Joseph Calbert & Heinz Vater (eds), 71–148. Tübingen: Narr.

Appendix

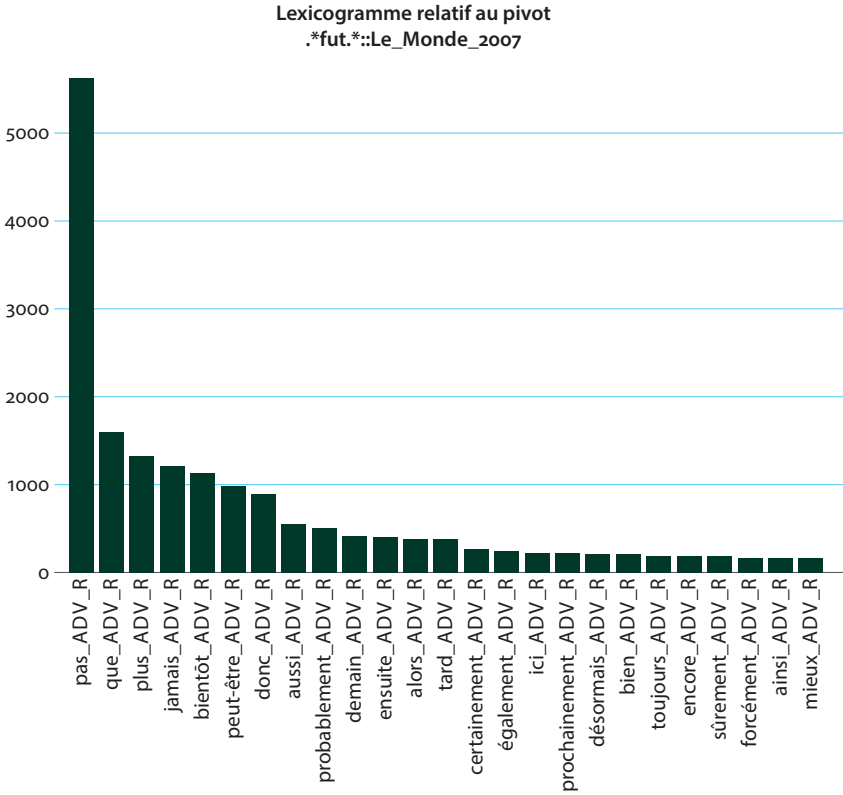


Figure A. Log-likelihood index of the 25 most salient adverbial collocates of the future tense in a right-context window of five words (source: <http://persan.rom.uni-koeln.de/btlsc/>)

The temporal uses of French *devoir* and Estonian *pidama* ('must')

Anu Treikelder and Marri Amon

University of Tartu

Using a parallel corpus, we compare the temporal uses of *devoir* in French and *pidama* ('must') in Estonian in order to identify possible correspondences in their future time reference. While they share similar properties in their root modality, they differ in their epistemic and postmodal uses. For French, we mostly follow the analysis of Hans Kronning (2001) who distinguishes three types of future-tense uses of *devoir*: alethic future, "subjective" and "objective" alethic future in the past. Our analysis demonstrates that unlike *devoir*, *pidama* does not have the "objective future in the past". In contrast, the data reveal a high degree of correspondence between the two verbs in the other future-related uses reported for *devoir* and generally absent in Estonian descriptions.

Keywords: modal verbs, postmodal uses, future, French, Estonian

1. Introduction

It is generally known that modality and tense, especially the future tense, are closely related categories from both the synchronic and the diachronic points of view. Future morphemes often develop from modal sources, as in the case of the French future, along with the future forms of other Romance languages, deriving from an obligative construction called *de-obligative* future by Dahl (2000). Additionally, from a synchronic perspective, future morphemes tend to convey modal meanings in addition to future time reference. In this chapter we take a different perspective: our purpose is to investigate how modal expressions can refer to the future.

We compare the postmodal and mainly temporal uses of modal verbs in two languages belonging to different language families: *devoir* in French and *pidama* ('must') in Estonian. Despite typological differences between these two languages, these two verbs seem to act quite similarly in the field of modality: they express root and epistemic necessity and both also have so-called postmodal uses that

cannot be described strictly in terms of necessity and are regarded as diachronic extensions of the primary modal meanings (cf. van der Auwera & Plungian 1998). However, whereas their usages are similar in the field of root modality, there are differences in their epistemic and especially postmodal values (cf. Treikelder 2015, 2016). This chapter will deal mainly with the expression of the future by means of these two verbs.

The term future is here used in a broad sense, including the cases where the event is situated in the future in respect to the moment of speech and also the cases where it is posterior to a moment of reference in the past (the so-called future in the past), since this label is given to both types of temporal uses of *devoir* in French (see Section 2 below).

The future value has been described as the main postmodal value of the French verb *devoir* (cf. Vetters 2004). Studies about the Estonian verb *pidama* have mostly mentioned three postmodal usages: quotative, avertive and intentional. The aim of this study is to find out whether there are correspondences between the postmodal uses in both languages, by focusing on the capacity of these two verbs to refer to the future. We comparatively examine the role of modal verbs in the creation of the future time reference of a situation, their interaction with other contextual elements and the relations of the temporal value with the primary modal meanings of these verbs. The contrastive perspective enables us to specify the properties of the situations, such as the intentional or non-intentional nature of the verb and the agentivity and the animacy of the subject, and, more generally, the contexts in which modal verbs receive a (dominantly) temporal reading.

The first part of the chapter is dedicated to the descriptions of the postmodal uses of the two verbs in previous studies in order to establish a primary framework for the comparison. The second part examines the correspondence of the two verbs in authentic translated examples drawn from a French-Estonian parallel corpus. For comparison, where available, we also added examples in English, that is, translations from Estonian or French into English. The examples are analyzed mainly from a qualitative point of view, given that translations do not always reflect linguistic nuances and may contain misinterpretations. However, the study of translational equivalents enables us to point out preferential choices of linguistic means as well as to test the possibility of using the corresponding modal verb in a particular context if a different solution appears in translation.¹ We have essentially taken a synchronic perspective, although some hypotheses on the grammaticalization of the two verbs can be drawn from the data.

1. See e.g. Celle (2006) for a discussion about using translations as data for language comparison.

2. The descriptions of the temporal uses of *devoir* and *pidama* in previous studies

This part of the study deals with the correspondences of the two verbs according to their descriptions in previous studies. We first report on some important disparities in the two languages before trying to find a common ground for comparison.

Neither of the two verbs studied has produced a future periphrasis used as a main future time reference device. In French, besides of a morphological future, the construction *aller* + infinitive has been grammaticalized in aspectual (prospective) and temporal uses. In Estonian, the grammatical expression of future is not obligatory. There is no flexional future, but different lexical means, including modal verbs, are used, mainly in the case of imperfective predicates (Metslang 1994: 148; cf. also Metslang 1996). In contemporary Estonian, the modal *pidama* does not figure among these devices although it has been used as a translational equivalent for the future in early texts (cf. Habicht 2001; Kilgi 2010). In a recent study on 17th-century Estonian, Habicht & Prillop (2016) identify the frequent use of *pidama* with the meaning of “predetermination”. They observe that sentences containing this use of *pidama* often are interpreted as conveying the meaning of “relative future” (Habicht & Prillop 2016: 51). The frequency of this use is attributed to the influence of the text type (translated religious texts) and considered as a “bridging context between deontic and epistemic modal meanings” (Habicht & Prillop 2016: 60).

However, the French *devoir* has been qualified as a future auxiliary under the process of grammaticalization. Kronning (1996: 16–18) makes an inventory of the labels given in different sources to the uses of the verb *devoir*. The label future figures in the description of several meanings (e.g. *futur proche* ‘near future’, *indéfini* ‘indefinite’, *probable* ‘probable’, *d’obligation* ‘future of obligation’). Additionally, other terms refer to future-related uses (e.g. *l’intention* ‘intention’, *le convenu* ‘agreed’, *la destination* ‘aim’, *la destinée* ‘destiny’). Picoche (1993: 166) gathers different uses of *devoir* under the term of *futur virtuel* (‘virtual future’) in which its basic meaning of necessity or obligation has become weaker and the future time reference prevails. Some authors make more precise distinctions between different future meanings. In our analysis, we have mostly relied on the perspectives of Kronning (1996, 2001) and Vetters & Barbet (2006).

Kronning (2001: 69) relates the future uses of *devoir* to his general description of this verb. He points to the need for distinguishing, besides the traditionally mentioned deontic (root) and epistemic modality of this verb, a third type that he calls “alethic modality”.² In this modality, *devoir* expresses “absolute necessity”, as in

2. Kronning has been criticized for using the term “alethic” conveying a different meaning in traditional logic (cf. Vetters & Barbet 2006). We use this term as it has been defined by Kronning.

the sentence *Si tu lances une pierre en l'air, elle doit retomber* ('If you throw a stone in the air, it must fall back down') (Kronning 2001: 68). This modal meaning thus covers certain usages of the verb *devoir* that are generally considered epistemic by others (e.g. Dendale 1994, 2000, cf. also Kronning 2001: 76).

According to Kronning (2001), both alethic and epistemic meanings pertain to the modality of being (*modalité de l'être*), whereas the deontic use of *devoir* belongs to the modality of doing (*modalité du faire*), requiring the agentivity of the subject of the modal verb and presuming that the action expressed by the infinitive is under the control of the subject (the participant-oriented modality in typological studies, cf. van der Auwera & Plungian 1998). The distinction between alethic and epistemic meanings is based on a different kind of properties reflected also in their syntactic and discursive behavior: the former involves an operation of veridiction, that is, the sentence can be judged in terms of truth or falseness, while the latter, expressing probability, does not allow this kind of operation. Kronning (2001) considers the future uses of *devoir* as subtypes of alethic modality, defining three types of future: alethic future with *devoir* in the present tense as in (1); subjective alethic future in the past as in (2) and objective alethic future in the past as in (3):

- (1) *Shimon Peres [...] est arrivé hier à Paris. Il doit s'entretenir ce matin à l'Elysée avec le président François Mitterrand, quelques heures avant l'intervention télévisée de ce dernier.* (Le Matin 17-08-1982: 9, 1; cited in Kronning 2001: 74)
'Shimon Peres [...] arrived to Paris yesterday. He is due to [must] meet with president François Mitterrand this morning at the Elysée, a few hours before the latter's appearance on TV'.
- (2) *Cependant, pour Lisbeth, la date du départ approchait; elle devait quitter Paris le dimanche suivant.*
(Martin du Gard 1922–1923: 225; cited in Kronning 2001: 78)
'However, for Lisbeth, the departure date was approaching; she was due to [must] leave Paris the next Sunday'.
- (3) *Le duc d'Aquitaine, Guillaume le Pieux, fonda au début du Xe siècle, à Cluny, [...] un monastère, d'où rayonna le prodigieux esprit clunisien dont l'influence civilisatrice sur le monde occidental devait être prépondérante.*
(Gimpel 1980: 9; cited in Kronning 2001: 78)
'The duke of Aquitaine, William the Pious, founded at the beginning of the 10th century in Cluny, [...] a monastery, from which spread the prodigious Cluniac spirit whose civilizing influence on the Western world was to [must] be predominant'.

These uses are firstly distinguished according to the temporal localization (present or past) of the reference point relative to which the situation is regarded as subsequent. The subjective and objective future are both viewed from a past reference point, but the latter implies that the action has really taken place, while the former means that the action is only expected, and it is not known whether it has been accomplished or not at the moment of speech (cf. also Nilsson-Ehle 1943–44). The interpretation of the objective future thus requires the existence of a supplementary viewpoint from which the realization of the action is verified.

Using the symbols introduced by Reichenbach (1947) and widely spread in tense and aspect studies (point of speech S, point of event E, point of reference R), the objective future receives the description: $R_1 - E - R_2, S$. This future is thus clearly distinct from other future in the past forms in French (i.e. the conditional form or the periphrastic future in the past), given that two points of reference are involved, and that the event is clearly anterior to the point of speech coinciding with the second point of reference.³ This is not the case for the other future in the past forms for which the temporal relation (anterior or posterior) to the point of speech remains unspecified. The objective future is usually considered the most grammaticalized future form of *devoir* in French.

Vetters and Barbet (2006) base their description of the future uses of *devoir* on the system of Kronning, challenging it to some extent, and propose a binary classification: (a) the future of destiny corresponding to the objective past use defined by Kronning (as in Example (3) above), and (b) the future of consensus (*le futur convenu*) englobing both the subjective past and the present uses (as in Examples (1) and (2) above). Vetters and Barbet (2006) consider that the second type of future cannot be considered alethic, in Kronning's terms, but should, rather, be associated with the root meanings of *devoir*. Unlike Kronning for whom all future uses of *devoir* are alethic, Vetters and Barbet divide them into two fundamentally different types. Unlike Kronning, they also regard them as postmodal uses, the first of which derives from the alethic modal meaning (modality of being) and the second from the deontic modal meaning (modality of doing). Table 1 below illustrates the correspondence between the two approaches.

3. Using these symbols to represent the other two types of future described by Kronning seems problematic. Less grammaticalized, these uses would receive a similar description to other "future in the past" forms in French. These schemes thus conceal some important semantic properties that *devoir* + infinitive in these future uses have in respect of other future devices in French and can be misleading.

Table 1. The future meanings of *devoir* according to Kronning (2001) and Vettters & Barbet (2006)

| Kronning (2001) | Vettters & Barbet (2006) |
|-----------------------------------------|-------------------------------------|
| Alethic future | |
| Example 1 | Future of “consensus” (not alethic) |
| “Subjective” alethic future in the past | Examples 1 and 2 |
| Example 2 | |
| “Objective” alethic future in the past | Future of “destiny” (alethic) |
| Example 3 | Example 3 |

In the descriptions of contemporary Estonian, the only postmodal value of *pidama* that has been directly associated with future reference is its intentional value, pointed out only for its preterit form (cf. Erelt 2001; Erelt & Metslang 2009). Erelt & Metslang (2009: 188) give an example illustrating the case where “intentional implication of agent-oriented necessity has developed into the main meaning of the verb *pidama*” in (4):

- (4) *Pärit oli ta Albuquerque'st [-], aga viimased 18*
 He came from Albuquerque [-], but he has lived the past eighteen
aastat elab Hawaiiil. Pidi algul minema 5 aastaks aga
 years in Hawaii. must:PST at_first go:MINF 5 year:TRNSL but
kuna neile naisega seal meeldis siis ei plaanigi
 because they:ALL wife:COM there please:PST:3SG then NEG plan:CLC
sealt ära tulla ilmselt
 thence VPTCL come.DINF apparently
 ‘He came from Albuquerque [-], but he has lived the past eighteen years in
 Hawaii. At first he intended to go there for five years, but as he and his wife
 liked it there, then apparently they have no plans to leave the place’
 (Google) (cited in Erelt & Metslang 2009: 188)⁴

Erelt (2001: 7) considers the intentional meaning as “a closely related meaning that has developed from the modal meaning”.

The other two postmodal values reported for *pidama*, that is, avertive and quotative, occur also only in the past and are highly grammaticalized forms. Following Bybee et al. (1991, 1994), Erelt and Metslang (2009) suggest that these uses have developed from the intentional value. The avertive meaning (was on the verge of V-ing but did not) contains an aspectual component of *imminence* (besides of the temporal *past* and modal *counterfactual* components (cf. Kuteva 2000)) and can

4. The glosses as well as the general translation are provided by Erelt and Metslang (2009: 188).

thus be related to the future to some extent. In contemporary French,⁵ this value does not exist for the verb *devoir* and the occurrences of avertive *pidama* are systematically translated with the verb *faillir* ('almost to do'), as it can be seen also in Example (5) derived from our corpus:⁶

- (5) a. “Reverdi n’a plus aucune chance d’échapper à la peine capitale. Son avocat s’est suicidé.” **Il faillit tomber de sa chaise.** *Le titre occupait la colonne de gauche du journal, sur la première page. Il lut seulement les quelques lignes qui introduisaient l’article.*
- b. “Reverdil pole enam mitte mingit lootust surmanuhtlusest pääseda. Ta advokaat sooritas enesetapu.” **Marc pidi** [must.PST.3SG] **toolilt maha kukkuma** [fall.MINF]. *Pealkiri seisis esilehel vasakus tulbas. Ta luges üksnes paari tutvustavat rida.* (FR-EST)
 “Reverdi has no chance any more to escape the capital punishment. His advocate has committed suicide.” He almost fell from his chair. The title occupied the left column of the newspaper, on the first page. He read only some lines that introduced the article.’

In the quotative meaning (= reported evidentiality: ‘allegedly’), the verb *pidama* in the preterit form can be used to situate reported actions in the past, the present (as in Example (6)), or the future in relation to the moment of speech. The quotative is thus not specifically related to any temporal reference. The French *devoir* does not have this use as shown in (6):

- (6) a. *Meile soovitati Genfis üht India restorani. Jube hea [awfully good] köök [cuisine] pidi* [must.PST.3SG] *olema* [be.MINF]. *Kallis küll aga...*
- b. *On nous a conseillé un petit restaurant indien à Genève... Il paraît que c’est délicieux. Pas donné, mais...* (EST-FRA)
 ‘We were recommended an Indian restaurant in Geneva. The food is said to be awfully good. Expensive but ...’

5. Interestingly, we can find the verb *faillir* (along with the periphrases *aller* + Inf and *être sur le point de* + Inf) in the translations of examples drawn by Veters and Barbet (2006) to illustrate one of the possibly temporal uses of the verb *devoir* in Old French. They consider *devoir* in these cases as an auxiliary of prospective aspect (the situation is “seen” from a moment preceding its beginning), thus expressing an aspectual rather than a temporal meaning.

6. Because of the length of corpus examples, a full interlinear morpheme-by-morpheme translation is not provided. The Estonian forms under study are indicated in square brackets directly in the text after the respective term. English translations come, where available, from the translated versions of the books.

There are thus quite remarkable disparities in the descriptions of *devoir* and *pidama* when it comes to the future time reference. Nevertheless, it seems that a certain common ground for comparison can be found in the intentional value reported for the Estonian *pidama*. The question is to what extent this value corresponds to the future values pointed out for *devoir*, whether other uses related to future reference can be reported for *pidama* and in what conditions they appear. Since the future reference of *devoir* has been more thoroughly described, the values of this verb served as a starting point for our corpus analysis.

3. The results of the corpus analysis

3.1 Data and method

Our comparative analysis is based on the Estonian-French parallel corpus CoPEF (cf. references) that contains fictional and non-fictional texts as well as includes translations in both directions. This corpus thus makes it possible to compare and analyze the use of linguistic means in the translations as well as in the original texts. The examples of the temporal uses of modal verbs were extracted manually from all the search results obtained for the French verb *devoir*. We classified the examples into three categories, thus adopting the tripartite system of Kronning (2001) (cf. Table 1 above) that distinguishes between past and present occurrences also in the case of the use that Vettters and Barbet (2006: 205) call “future of consensus”. This distinction was important from the comparative point of view, since the only future-like usage that has been described for the Estonian verb *pidama* (i.e. the intentional use) has been associated exclusively with its past forms. However, we shall also refer to Vettters and Barbet’s description in our analysis. The analysis is essentially qualitative because number of examples is not very large (cf. Table 2 below). This approach also results from the nature of our corpus, given that translation equivalents do not always reflect an actual linguistic correspondence between the two verbs. It is mostly possible to use the correspondent verb in translations but, in some cases, it changes the meaning of the sentence.

We will examine possible semantic shifts when the verb *pidama* is used as well as test the possibility of using it when different means occur in the Estonian sentence. The identification of the future uses in authentic data is not unproblematic. Modal verbs as polysemous markers frequently have more than one interpretation, even if the wider context is taken into account, and it is not possible or necessary, from the communicative point of view, to draw a distinct line between different values. Obviously, all the cases where the lexical verb in the infinitive refers to a subsequent situation cannot be considered as future uses in the narrow postmodal

sense. Obligation in a present context is generally oriented to the future and epistemic evaluation may concern an expected situation as well. For the French verb *devoir*, different tests have been proposed to distinguish its future uses from deontic and epistemic meanings.

In order to distinguish between the deontic and the temporal use of the verb *devoir*, Picoche (1993: 38) uses the pronominalization test, mentioned also by Vetters and Barbet (2006). This test indicates the impossibility of replacing the infinitive complement with a personal pronoun in the future uses (*Daniel doit aller au cinéma ce soir* ‘Daniel has to go to the cinema tonight’ → **il le doit*), like in the epistemic uses of *devoir*. On the other hand, different paraphrases are proposed to distinguish between the values of *devoir*. For its deontic value, Picoche (1993: 38) mentions “*c’est son devoir*” (‘it is his obligation’), “*il est bien obligé*” (‘he doesn’t have a choice’) or “*il ne peut pas faire autrement*” (‘he can’t do otherwise’) that do not apply in the case of the future use of *devoir*, paraphrased rather by “*il a l’intention, le projet de ...*” (‘he has the intention to, he plans to ...’). Kronning (2001: 68) points out the impossibility of paraphrasing the modal construction *devoir* + Inf with an epistemic adverb of the type *probablement* (‘probably’) in the case of its alethic value (including the future value) in contrast with its epistemic value. He also observes that the alethic *devoir* cannot be repeated by the anaphoric nominal phrase *cette obligation* (‘this obligation’). This property distinguishes alethic and deontic *devoir*, the latter allowing this kind of anaphora. Additionally, Kronning (2001: 73) points out specificities in the syntactic and discursive behavior of the alethic *devoir*, especially its compatibility with *wh*-questions and causal subordinate clauses introduced by *puisque* (‘since’). In this respect, the alethic *devoir* acts differently compared to the epistemic *devoir* but, as Vetters and Barbet (2006) indicate, it shares this property with the deontic use.

However, as Picoche (1993: 38) also observes, tests do not always allow one to determine one particular value while ruling out the others. In our corpus, this is especially true in the case of alethic future (with the modal verb in the present tense) and subjective future in the past, since these meanings remain close to the primary modal meanings of *devoir*. We have excluded highly ambivalent cases and picked out only the examples where the future reading seemed the most plausible one. Some difficulties arose also in distinguishing between the objective and the subjective future in the past. We shall discuss these cases later.

Additionally, we have excluded from our corpus two types of examples where *devoir* + Inf refers to the future. We do not consider the so-called generic sentences that refer to a generally acknowledged fact, in examples such as *tout être humain doit mourir un jour* (‘every human being must die one day’). Kronning (1996, 2001) analyses similar examples under a type of alethic *devoir* (expressing “synthetic necessity”, i.e. law of nature), different from the alethic future (their

differences are thoroughly discussed in Kronning 1996: 70ss.). In comparison, the Estonian *pidama* does not seem to have significant differences in this respect, being able to convey this meaning like necessity modal verbs in many other languages (cf. Kronning 1996: 94). The second type of examples that we excluded concern the use of *devoir* in conditional subordinate clauses after *si* ('if'). The verb *pidama* occurs in these cases generally in the conditional mood, since Estonian allows the use of the conditional form after the conjunction *kui* ('if'), but it does not have a flexional future form. These cases thus involve several morphological and syntactic peculiarities in the compared languages that cannot be discussed here and deserve a separate study.

After excluding the examples described above and the ambiguous cases with a strong possibility of modal reading (deontic or epistemic), we had retained 139 examples for our comparative analysis. The distribution of the examples between the future meanings and translation directions as well as the rate of *pidama-devoir* correspondences are displayed in Table 2.

Table 2. Corpus examples according to different future meanings and *pidama-devoir* correspondence

| Future meaning | Number of occurrences of <i>devoir</i> | | Total number | Correspondence <i>devoir-pidama</i> (%) | | Total correspondence (%) |
|-------------------------------|----------------------------------------|--------|--------------|-----------------------------------------|----------|--------------------------|
| | FR-EST | EST-FR | | FR-EST | EST-FR | |
| Alethic future | 26 | 19 | 45 | 16 (62%) | 10 (53%) | 26 (58%) |
| Subjective future in the past | 21 | 26 | 47 | 18 (86%) | 23 (88%) | 42 (89%) |
| Objective future in the past | 31 | 16 | 47 | 6 (19%) | 0 | 6 (13%) |
| Total | 78 | 61 | 139 | 40 (51%) | 33 (54%) | 74 (53%) |

The corpus includes examples from 23 French authors and 17 Estonian translators as well as 9 Estonian authors and 7 French translators. Although the number of authors and translators from Estonian to French is smaller, we believe that the corpus has sufficient variability for both languages. When it comes to variation between the two directions of translation, Table 2 reveals no significant differences in the *devoir-pidama* correspondence rate for the first two categories. The divergence is more noteworthy in the case of the third category, that is the objective future in the past, which appears much more often in French original texts than in translations (31 vs. 16 occurrences) and reveals a *devoir-pidama* correspondence only in French-Estonian translations. This difference will be discussed further in the Section 3.2.

3.2 The “objective” future in the past

The results of the analysis show that in the case of the objective future in the past (the future of destiny in Vetter & Barbet 2006) the correspondence rate between *devoir* and *pidama* is very low (13%, cf. Table 2). Generally, this type of sentence is translated in Estonian with the main verb in the past and a temporal adverb, as in Example (7):

- (7) a. *Comment ma mère, si scrupuleuse, si attentive, et dont l'inquiète sollicitude me devait même bientôt excéder, comment sa vigilance ici s'endormait-elle?*
 b. *Kuidas võis mu ema, kes muidu oli ülearu ettevaatlik ja tähelepanelik ning kelle muretsevad hoolitsused mulle õige varsti [soon] talumatuks [intolerable] muutusid [become.PST.3PL], sel korral valvsuse minetada?*
 c. How was it that my mother, who was usually so scrupulous, so attentive, whose anxious care I soon **found** actually **irksome**, how was it that my mother allowed her vigilance to flag upon this occasion? (Gide 1915: 52, FR-EST)

According to Kronning (2001: 78), the interpretation of the verb *devoir* in this future use involves two viewpoints: a subjective point of view in the past from which the situation is regarded as forthcoming and a second, objective point of view from which the situation is confirmed to have actually taken place. This second, subsequent viewpoint is clearly dominant in these uses of *devoir*, as it gives the final interpretation and the actual meaning to the construction and it temporally corresponds to the moment of speech. The Estonian preterit verb form in Example (7b) merely expresses this retrospective validation of the situation, while the predictive meaning is conveyed by the adverb *varsti* ('soon'). As it can be seen, the simple past tense is also used in the English version in (7c). This use of *devoir* is known also as future of fate or future of destiny (cf. Vetter & Barbet 2006), since it often expresses the inevitable nature of the situation, as in Example (8):

- (8) a. *La mort prématurée de son fils Octave, entré tout jeune à l'École polytechnique et dont il voulait faire un chef porta un coup terrible à Olivier Blévine. Il ne devait pas s'en relever et mourut deux ans plus tard en février 1908.*
 b. *Kui tema poeg Octave, kes oli väga noorena Pariisi Kõrgemasse Polütehnilisse Kooli astunud, et juhtimine enda kätte võtta, varakult suri, siis andis see isale valusa hoobi, millest ta [PRON] ei [NEG] toibunudki [recover.PST.3SG]. Isa lahkus aasta hiljem, veebruaris 1908.*
 c. The premature death of his son Octave, who had entered the Ecole Polytechnique at a very early age and of whom he wanted to 'make a leader' was a terrible blow to Olivier Blevigne. **He was never to recover** from it and **died two years later, in February, 1908.** (Sartre 1964: 48, FR-EST)

The predictive viewpoint can be expressed lexically by the Estonian verb, as in Example (9) where the passive verb form *olema määratud* ('be destined to') also conveys the fatality of the situation. This example comes from an Estonian original text:

- (9) a. *Nõnda et härra erupolkovnik von Bock oli tüürimas militaria'st litteraria'sse. Ja minust plaanitses tema endale kirjutajat. Aga seesinane tema kava [plan] oli määratud [was destined] veel vähem [even less] täide minema [come to fruition] kui minu plaan valitsejakohast Võisiku.*
- b. *Ainsi donc le colonel en retraite von Bock était-il en train de naviguer des militaria vers les litteraria. Et de moi il envisageait de faire son copiste. Mais ce plan devait encore moins se réaliser que celui conçu par moi d'exercer à Võisiku les fonctions d'intendant.*
- c. Thus, retired colonel von Bock was contemplating a move from the military to the literary realm, with me as his scribe. But the plan was **destined to come** even less **to fruition** than my own intention to become steward of Võisiku. (Kross 2003: 29, EST-FR)

Whereas most of the examples of this future meaning contain a temporal adverb in our corpus, the Examples (8) and (9) above show that the use of an adverb is not always necessary. In some cases, adverbials appear only in Estonian texts. It can thus be assumed that *devoir* + Inf combined with the meaning of the infinitive verbal form by itself contains the necessary predictive viewpoint, while in Estonian it has to be marked lexically.

The fact that a certain event has really taken place (retrospective assessment) can be deduced from different elements, also depending on the type of text. For instance, in non-fiction texts some examples refer to real historical events that have taken place afterwards. Diary-like storytelling also often contains clear elements that refer to knowledge acquired later, such as references to later written sources or discoveries made subsequently. In fictional texts, the references can be also extratextual or historical, but mostly the retrospective viewpoint is contained in the context, either explicitly (Example 8) or implicitly (Examples 7 and 9). The verb *devoir* can thus also convey this reading in the cases where no other elements in the immediate context overtly refer to the realization of the action.

Based on the analysis, one can say that the Estonian modal verb *pidama* is not an appropriate equivalent for the French objective alethic future uses of *devoir*, as it has been pointed out also by Amon (2016). Nevertheless, our corpus contains six examples (cf. Table 2) of the verb *pidama*, all in translations from French into Estonian. Some of these examples appear to be simple misinterpretations, but there are cases where the use of *pidama* seems appropriate or at least possible as in (10):

- (10) a. *Ce pot de départ à la retraite devait constituer le dérisoire apogée de mes relations avec le ministère de l'Agriculture. J'avais recueilli tous les renseignements nécessaires pour préparer mes cours; nous n'aurions plus guère à nous revoir; il me restait une semaine avant de partir à Rouen.*
- b. *See koosviibimine [this party] pidigi [must.PST.3SG:CLC] jääma [stay.MINF] minu ja põllumajandusministeeriumi vaheliste suhete naeruväärseks kõrghetkeks. Ma olin saanud kogu vajaliku teabe selleks, et ette valmistada koolituskursus; meil pole enam mingit põhjust kohtuda; Roueni sõiduni jäi veel üks nädal.* (FR-EST)
 'This farewell party was destined to be the ridiculous climax of my relationship with the Ministry of Agriculture [...]'

In (10b), *pidama* + Inf without a temporal adverb seems to have a meaning very close to the future of destiny, reinforced by the use of the enclitic particle *-gi*. This particle has been associated with expectations or predictability (Metslang 2002); here it contributes to stressing the predestined consequence of the event. A similar use of *pidama* occurs also in original texts, as in (11a), which is not included in our corpus since the verb *devoir* does not appear in the French version. In French, the periphrastic future *aller* + Inf is used with the verb *tarder* conveying the meaning of the adverb *varsti* ('soon'):

- (11) a. *Kuid ma tahan siia meelespidamiseks mõned Timo sõnad sellest kevadest kirja panna, mis varsti [soon] suurema tähenduse pidid [must.PST.3PL] omandama [assume.MINF], kui neil esialgu paistis olevat.*
- b. *Je voudrais cependant noter ici pour mémoire quelques paroles prononcées par Timo ce printemps-là, des paroles qui n'allaient pas tarder à revêtir une signification plus grande que celle qu'elles semblaient avoir au début.*
- c. Here, I would like to record some of the things Timo said that spring; they were soon to assume a greater significance than it seemed at first.
 (Kross 2003: 51, EST-FR)

Although these examples seem to offer a similar reading for *pidama*, they cannot be entirely assimilated into the objective future use of *devoir*. First, they are rather marginal among the uses of this verb, but it seems also possible to connect these cases of *pidama* to its subjective future uses (cf. 3.3 below). In this case, the modal verb *pidama* would only be responsible for the predictive viewpoint in these examples (along with the adverb *varsti* 'soon' in Example (11)), and the retrospective viewpoint is provided by other contextual elements (besides the general past context, e.g. the verb *jääma* 'stay' in (10) and *kui neil esialgu paistis olevat* 'than it seemed at first' in (11)). A future study, taking the verb *pidama* as its starting point, should be conducted to find out the extent and precise contexts of this use.

We conclude that, in general, the verb *pidama* does not have an objective reading, but it can appear as a translational equivalent of *devoir* in certain contexts where the internal subjective viewpoint is available. Whereas the uses of the objective future can generally be quite easily identified, in this type of text the difference between the objective and subjective readings is not always clear-cut. Even if it is known from the context that the event has taken place, the subjective internal viewpoint may prevail or can be considered as an alternative reading.

3.3 The subjective future in the past

The subjective future in the past (the future of consensus in Veters & Barbet 2006) reveals, in contrast, a very high rate of *devoir-pidama* correspondence (89%). One of these cases is illustrated in Example (12):

- (12) a. *See oli üsna lahe noorhärna ja austusest Timo vastu kohtles ta Eevat kogni kavalerlikult. Kui ta saabus, pidi [must.pst.3sg] ta jääma [stay.minf] Võisikule pikemaks [for longer], aga nädala pärast sõitis ta äkitselt ära.*
- b. *C'était un jeune homme d'un commerce plutôt agréable et qui, par respect envers Timo, se comporta envers Eeva de la manière la plus courtoise. Quand il arriva, il devait rester quelque temps à Võisiku, mais au bout d'une semaine, brusquement, il repartit.*
- c. He was a very pleasant young gentleman, and out of respect for Timo he treated Eeva as though he were a gallant admirer. He came to us **with the intention of staying** at Võisiku *for an extended period, but left abruptly after only a week.* (Kross 2003: 51, EST-FR)

Example (12) is an unambiguous case of the subjective alethic future in the past. In Kronning's (2001: 78) terms, this future meaning presumes only a predictive viewpoint in the past and nothing indicates whether, at the moment of speech, the action has been accomplished or not. In (12), the predictive viewpoint is localized by the clause *kui ta saabus/quand il arriva* ('when he arrived'). It is indicated in the immediate context that the action did not take place (*but left abruptly after only one week*), but this meaning is not conveyed by the modal verb. As we will see below, there are cases of subjective reading in which it is known that the action was actually completed. This information is thus not necessary for the subjective future reading of the verb *devoir*. Veters and Barbet (2006) argue that this use of *devoir* (along with its use in the present, cf. 3.4 below) is not alethic, but should rather be associated with its main modal value of obligation. As Veters & Barbet (2006) point out, sentences of this kind convey the meaning of an agreement or a consensus (cf. future of consensus), according to which the action is expected to

take place, that can be interpreted as a sort of obligation in a weak sense for the person involved (cf. also Kronning 1996: 130 for the relations between deontic and alethic future interpretations). Nevertheless, this use cannot be regarded as a subtype of root modality, since it does not allow the pronominalization of the infinitive verb (cf. Vetters & Barbet 2006 and 3.1 above). Example (12) above may indeed be interpreted in this way: we can presume that there is a sort of an agreement between the young gentleman and his hosts concerning his stay at Võisiku. On the other hand, this example looks very much like Example (4) provided above for the intentional value of *pidama*. Example (12) also contains an intentional predicate controlled by the subject of the sentence. In the English version, the intentional meaning is clearly dominant (*with the intention to*). We can thus assume that the French conventional future and the Estonian intentional use can be assimilated to a certain extent. However, there are instances where only the value of consensus seems to appear as a possible interpretation. In Example (13), the subject of the sentence in French and in Estonian is inanimate (*le rassemblement/kogunemine* ‘gathering’) and the verb is non-intentional (*se faire/toimuma* ‘take place’), but *pidama* is still an appropriate equivalent for *devoir*:

- (13) a. *Ainsi s'organisait lentement l'expédition [...]. Le rassemblement devait se faire un jour de juin 1938, aux portes de la ville, d'où bœufs et cavaliers se mettraient en route sous la direction de Fulgencio, avec une partie des bagages.*
- b. *Nii valmistasime ekspeditsiooni ette tasapisi [...]. Kogunemine [gathering] **pidi** [must.PST.3SG] **toimuma** [take_place.MINF] ühel 1938. aasta juunikuu päeval linnavärvate juures, kust härjad ja ratsanikud **pidid** Fulgencio juhitud osas pagasiga teele asuma.*
- c. And so, slowly, [...] our expedition got organized [...]. We were due to assemble one day in June 1938 at the gates of Cuiaba, whence oxen and riders were to set off to rendezvous with Fulgencio, taking some of our baggage with them. (Lévi-Strauss 1961: 254, FR-EST)

Several examples of subjective *devoir* and *pidama* in our corpus seem to allow neither consensual nor intentional reading, as in (14):

- (14) a. *Jumal ise teab, kust keiser selle võttis, sest Timo ja Eeva laps **pidi** [must.PST.3SG] ju [PRTCL] alles viie kuu pärast [only after five months] **sündima** [be born.MINF]!*
- b. *Dieu seul sait où l'empereur était allé chercher cela: l'enfant d'Eeva et de Timo ne devait naître que cinq mois plus tard!*
- c. God knows who had told the Czar this: Timo's and Eeva's child was not due for another five months. (Kross 2003: 68, EST-FR)

In Example (14), the infinitive *sündima/naitre* ('be born') is a non-intentional verb and the situation expressed in the sentence does not involve any agreement between the subject and somebody else. This sentence, rather, conveys the meaning that the event (*the birth of the child*) is expected to happen. Thus, here we have a different type of the subjective future that seems to be related not to a weakened obligation (i.e. root modality) but to the knowledge about the situation, i.e. epistemic modality. However, this use cannot be assimilated to the epistemic use of *devoir/pidama*, because it is not possible to replace the modal verb by an adverb (**l'enfant ne naitrait probablement que cinq mois plus tard*, cf. Kronning 2001 and 3.1 above). A similar case is illustrated in Example (15), but the subject is inanimate and the modal verbs seem to additionally receive a fatal meaning conveyed by *devoir* in its objective uses:

- (15) a. *Pealegi oli Mari juba eelmisel ööl mingit halba und näinud. Halb uni ei või kunagi head tähendada, arvas Mari, ja sellepärast ta aina ootas [waited], et millal see kätte tuleb [when this would happen], mis [what] tulema [come. MINF] peab [must.PRS.3SG].⁷*
- b. *De plus, elle avait fait un mauvais rêve la nuit précédente, ce qui ne pouvait selon elle rien présager de bon, aussi attendait-elle que ce qui devait arriver lui tombe dessus.*
- c. Mari had a bad dream, which never boded well, and after she woke, she waited to see **what would happen next**. (Tammsaare 2014: 277, EST-FR)

Moreover, compared with Example (12), no information is provided in the immediate context about the accomplishment of the action in Examples (13) through (15). Even if it appears later in the text that the event has really taken place (the gathering in (13) and the childbirth in (14)), these examples provide a clearly subjective reading. So, although there might be a possible ambiguity between subjective and objective readings of whether or not the action has actually been completed, only the predictive viewpoint is available in these cases: the focus is on the moment in the past when this event was planned or expected and not on the retrospective validation of the situation. This possible ambiguity seems to concern mainly the French sentences with *devoir* in (14) and (15), as this verb more naturally receives an objective reading. We argue that the interpretation of the Estonian sentences in these examples requires only one (predictive) viewpoint. Similarly to the objective future, the examples of this value usually contain a temporal adverb, but this is not mandatory: the temporal reference of the expected situation may remain indefinite (cf. Example 15). The adverb often makes it possible to distinguish the future uses from the modal uses (root or epistemic).

7. The modal verb *pidama* is in the present because the Estonian does not have the constraint of sequence of tenses.

The examples provided above reveal that the Estonian *pidama* is an appropriate equivalent for *devoir* in its different uses of subjective future in the past. In the cases where other linguistic means appear in Estonian texts (5 occurrences), it would be possible to use *pidama*, without a significant change in meaning. The Estonian verb *pidama* in the preterit form thus seems to have other types of future-like uses besides its intentional meaning reported in previous studies.

In his study about the typology of future grams, Dahl (2000) distinguishes between the intention-based and the prediction-based future time reference. This distinction has to do with the animacy of the subject and the subject's control over the action. Dahl (2000: 310) points out that intentions are "by definition under our control (or at least believed to be so) and prototypically show up in sentences with a human subject (who is also the bearer of intentions)". In this regard, Example (12) represents a prototypical case of intention.

Predictions, on the other hand, "concern courses of events that are not within human control or at least not within the control of the speaker" (Dahl: *Ibid.*). Examples (14) and (15) can thus be qualified as predictions in this sense. In contrast, Example (13), although the subject (*le rassemblement* 'gathering') of the sentence is inanimate, cannot be regarded as a prediction since human control is implied in the event (humans have made the agreement to gather). It thus seems that this consensual meaning can be considered a non-prototypical case of intention or as an intermediate stage in the development from the intentional meaning into the predictive meaning. Dahl (*Ibid.*) observes that "markers that are originally restricted to intention-based FTR [future time reference] tend to develop into general future markers, which include prediction-based FTR as central cases but can in the normal case still be used for intention-based FTR". Examples (14) and (15) thus present more grammaticalized future uses of the modal verbs than Examples (12) and (13), but the latter has some common features with the more grammaticalized examples. We can hence suggest that the three meanings represent a progress on the scale of grammaticalization: intention > consensus > prediction. It is, however, important to remember that neither of the modal verbs (*devoir/pidama*) has become a grammatical future device. Besides the future reference, both convey an additional meaning ('be destined to') also in the most grammaticalized use (prediction).

The analysis of the corpus examples also seems to indicate that, whereas it is well justified to make the distinction proposed by Vettters and Barbet (2006) between the consensual meaning (pertaining to the modality of doing), and the alethic meaning (pertaining to the modality of being), this division does not coincide with the distinction between the "objective" and "subjective future in the past" proposed by Kronning (see Table 1 above). The examples provided in Section 3.2 indicate that the objective future uses, indeed, always express the modality of being: the sentences contain non-agentive (often but not exclusively inanimate) subjects and the event is not under the control of the subject. The subjective future uses,

however, are not restricted to a “consensual” use (close to the deontic modality), but can also pertain to the modality of being (and are alethic in Kronning’s sense). We have also pointed to a possible ambiguity in these cases between the objective and the subjective reading in French.

From our comparative perspective, we can conclude that *devoir* and *pidama* behave in a very similar way in respect to the “subjective future in the past” meaning.

3.4 The “alethic future”

According to Kronning (2001), this type of future includes the cases where *devoir* is in the present tense and the expected situation is thus located in the future in relation to the moment of speech. Veters and Barbet (2006) view this type as a “future of consensus” together with the subjective future in the past. Our corpus reveals, however, a much lower rate of *devoir-pidama* correspondence in this case (58% vs. 89% for the subjective future in the past). This value is the most problematic one from our comparative perspective. First, the examples of *devoir* in this use are not numerous, probably because of the nature of our corpus (mainly fictional texts). They mostly occur in dialogues or in diary-like narration. We have thus included all the examples found in the parallel corpus in our analysis. There is also more variation in translations and it is often difficult to clearly distinguish these occurrences from the primary modal uses. The interpretation depends on the nature of the situation expressed by the infinitive and the primary modal meanings always seem to be more or less present according to the verb type. Typically, the future-dominant interpretation occurs with the main verb (mostly a motion verb) expressing a perfective action, like in Examples (16) and (17):

- (16) a. *Peidupaik kolmele hingele on laadungi sees valmis. Toidu- ja joogivaru on nende jaoks paigas. Nad ise peavad [must.PRS.3PL] homme õhtupoolikul [tomorrow in the evening] siia pärale jõudma [arrive.MINF].*
- b. *La cachette ménagée pour trois personnes dans la soute est prête. Les vivres nécessaires y sont en place. Tous trois doivent arriver ici demain dans la soirée.*
- c. Now the new cargo has been stowed in the hold with a hiding place and provisions for three persons inside it. The three **are expected** here *tomorrow afternoon*. (Kross 2003: 269, EST-FR)
- (17) a. *Du moins, une chose est sûre, j’attends le missionnaire qui doit venir me remplacer.*
- b. *Üks asi on vähemalt kindel, ma ootan misjonäri, kes [who] peab [must.PRS.3SG] tulema [arrive.MINF] minu asemele.*

- c. At least one thing is certain, I am waiting for the missionary **who is to come** and take my place. (Camus 1957: 13, FR-EST)

Examples (16) and (17) illustrate the correspondence between *devoir* and *pidama*. We can thus see that, although in previous studies, intentional value has not been reported for the verb *pidama* in the present tense, it can occur with this meaning. In these examples, both consensual and intentional readings seem possible. The temporal adverb is generally present (like in Example (16)), but the time reference may remain unspecified (like in Example (17)). Example (18) shows that *pidama* and *devoir* may also correspond if the subject is inanimate and a fatal meaning is conveyed (cf. also Example (15) above):

- (18) a. [...] *ja korraga teab poeg, et nõnda see* [this] *peab* [must.PRS.3SG] *sündimagi* [happen.MINF], *et nüüd pole kindlasti mingit pääsu enam...*
 b. *Et, subitement, le fils comprend que cela doit se passer ainsi, que, maintenant, il n'y a absolument plus d'espoir...* (EST-FR)
 'And suddenly the son knows that this is going to happen, there is definitely no more hope.'

However, in several examples different means are used in the Estonian sentence. It seems that, even if it might be possible to use the verb *pidama*, this is not the most natural choice. For instance, in Example (19), the expression *on tagasi oodata* ('is expected back') is used, while a construction with *pidama* would also be possible:

- (19) a. *Paistab, et palju rohkem valmistab talle muret ema, keda* [who] *on* [be.PRS.3SG] *Moskvast tagasi* [back] *oodata* [expect.DINF] *õhtul*.
 b. *Il paraît beaucoup plus préoccupé par sa mère, qui doit rentrer de Moscou dans la soirée.* (EST-FR)
 'He seems much more worried about her mother who is expected back from Moscow in the evening.'

It is also noteworthy that in some examples, the verb *pidama* is in the past tense, while the event is expected in the future at the moment of speech:

- (20) a. *Mais il faudrait partir. On doit me téléphoner chez moi et puis, on nous regarde beaucoup ici.*
 b. *Aga me peaksime minema hakkama. Mulle* [PRON1SG.ALL] *pidi* [must.PST.3SG] *koju* [at home] *helistatama* [call.MINF.IMPS], *ja pealegi vahitakse meid siin liiga palju.* (FR-EST)
 'But we should go. I'm expecting a call and besides, people are staring at us here.'

In (20), *pidama*, like *devoir*, conveys the meaning of an agreement that has been made previously. This example, unique in our corpus, contains an interesting construction in Estonian, with an impersonal infinitive form (marked by the affix *-ta-*) and no subject, but as in the case of the impersonal voice in general, as well as for the French pronominal subject *on*, a human agent is implied and the intentional or consensual reading is possible. The past tense seems to refer to the agreement made before the moment of speech. The next example indicates that this use is possible also for the past form of *devoir*:

- (21) a. *Emä (naeratades): Ma pakun midagi juua äkki? Sõbranna: Kuule üks aperitiiv kuluks ära küll. (Vaatab kella.) Kuidas kell lendab täna! Juba pool seitse. Pierre pidi [must.PST.3SG] mulle seitse järgi tulema [come.MINF].*
- b. *La Mère (avec un sourire): Je vous offre à boire? L'Amie: Oui, je prendrais bien un petit apéro. (Elle regarde l'heure.) On ne voit pas le temps passer aujourd'hui! Déjà six heures et demie. Pierre devait venir me chercher à sept heures. (EST-FR)*
 '[...] (Looking at her watch) How the time flies today! Six thirty already! Pierre promised to pick me up at seven.'

However, we have not detected other examples of this kind in our corpus.

Based on our analysis, it seems that *pidama* in the present tense also has future-like uses, but they are less frequent than *devoir* in these contexts. In Estonian originals and in translations, other means are often preferred to the modal verb. The verb *pidama* in present tense is thus clearly less grammaticalized than its preterit form, but it also has postmodal temporal uses not reported before. However, the number of examples in our corpus is not sufficient to properly assess the possibilities of using the present form of *pidama* in this context. Additionally, we note that this future use also provides examples of both intentional/consensual and predictive meanings of *devoir* that were pointed out for the subjective future in the past.

4. Conclusion

The aim of this chapter was to use a translational corpus to determine whether similarities could be found in the future time reference of the modal verbs of necessity *devoir* and *pidama* ('must'). Whereas considerable differences appear in the descriptions of their postmodal values, our study has revealed that there are quite frequent correspondences between these two verbs in certain types of future-related uses reported for the French *devoir*.

Following Kronning (2001), we separately analyzed three types of future-uses. The most remarkable differences between the two verbs concern the "objective

future in the past”, conveying the meaning that something that was predicted to happen actually took place. The analysis of translations shows that *pidama* does not have this temporal use. Some marginal examples in the corpus where *pidama* can be considered as an adequate equivalent for *devoir* have to be classified instead among its subjective future uses: in these examples, the modal verb seems to be responsible only for the predictive viewpoint, the retrospective validation of the event being provided by contextual elements.

In contrast, our analysis reveals a great resemblance between the two verbs in the other future-related uses of *devoir* pointed out by Kronning (2001), that is, the “subjective future in the past” and the “alethic future”, both regarded as the “future of consensus” by Vetters & Barbet (2006). The correspondence in translations is particularly great between the past forms of the modal verbs and *pidama* seems acceptable in all examples of this use in our corpus. As concerns the “alethic future” with *devoir/pidama* in the present tense, this use seems to be less grammaticalized in Estonian, since the correspondence between the two verbs is not very important. Moreover, due to the particularity of the corpus that consists mostly of fictional texts, the present tense was overall less frequent. However, the present form of the verb *pidama*, like its past form, occurs in contexts where *devoir* is used for referring to the future from a subjective point of view: besides the intentional value, the future reference also encompasses the cases with “consensual” meaning as well as a type of future where the stress is on the subjective expectation of a future event. It seems that the latter can be considered as “alethic” in the meaning used by Kronning, and is thus related to the modality of being, like the epistemic uses of the modal verbs, since this future use occurs in sentences without a human subject and the event in the infinitive is not under the control of the subject (nor can the controlling subject be inferred from the context like for some “consensual” uses). We have thus argued that the alethic use is not restricted only to the “objective future in the past”, as suggested by Vetters and Barbet 2006 (cf. Table 1), but that it appears also in the other two types of future proposed by Kronning. The consensual meaning pointed out by Vetters and Barbet is less grammaticalized and occurs only in the case of one subjective viewpoint involved in the past or in present.

According to our analysis, *pidama* thus has a wider range of future-related uses than the intentional use in the preterit form previously reported for this verb. The correspondences between *devoir* and *pidama* in the corpus mostly concern the future-related uses with a relatively low degree of grammaticalization. In comparison with the “objective future”, they stay closer to the primary meanings of the modal verbs, especially to the meaning labelled as “participant-external non-deontic necessity” (van der Auwera & Plungian 1998: 82). The most grammaticalized postmodal uses of these two verbs, i.e. the “objective future” of *devoir* and the avertive and quotative uses of *pidama*, display very different semantic properties

in their actual uses. They seem thus to have taken different paths from a relatively similar modal starting point. These paths could not be investigated in depth in the framework of the present study which focused on the synchronic comparison of the functioning of the two modal verbs in future contexts.

Acknowledgements

We are grateful to the anonymous reviewers for their useful remarks and comments on the first version of this chapter. All errors are our own.

Abbreviations

| | | | |
|------|-----------------------|-------|-----------------------|
| 1 | 1st person | MINF | <i>ma</i> -infinitive |
| 2 | 2nd person | NEG | negation |
| 3 | 3rd person | PL | plural |
| ADE | adessive | PRON | pronoun |
| ALL | allative | PRS | present |
| CLC | clitic | PST | past |
| COM | comitative | PTCL | particle |
| COND | conditional | PTCP | participle |
| dINF | <i>da</i> -infinitive | QUOT | quotative |
| EST | Estonian | SG | singular |
| FR | French | TRNSL | translative |
| IMPS | impersonal | VPTCL | verb particle |

References

- Amon, Marri. 2016. Prantsuse modaalverbi *devoir* temporaalse kasutuse tõlkimisest eesti keelde (About translating temporal uses of the French modal verb *devoir* into Estonian). *Eesti ja soome-ugri keeleteaduse ajakiri/Journal of Estonian and Finno-Ugric Linguistics* 7(2): 13–26.
- Bybee, Joan L., Pagliuca, William & Perkins, Revere D. 1991. Back to the future. In *Approaches to Grammaticalization*, Vol. 2 [Typological Studies in Language 19: 2], Elizabeth C. Traugott & Bernd Heine (eds), 17–58. Amsterdam: John Benjamins. doi:10.1075/tsl.19.2.04byb
- Bybee, Joan L., Pagliuca, William & Perkins, Revere D. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago IL: The University of Chicago Press.
- Camus, Albert. 1957. *Exile and the Kingdom*, trans. by Justin O'Brien. New York NY: Random House.
- Celle, Agnès. 2006. *Temps et modalité: l'anglais, le français et l'allemand en contraste*. Bern: Peter Lang.

- CoPEF. Corpus parallèle estonien-français de l'Association franco-estonienne de lexicographie.
<<http://corpus.estfra.ee>>
- Dahl, Östen. 2000. The grammar of future time reference in European languages. In *Tense and Aspect in the Languages of Europe*, Östen Dahl (ed.), 309–328. Berlin: De Gruyter.
doi:10.1515/9783110197099.2.309
- Dendale, Patrick. 1994. *Devoir* épistémique, marqueur modal ou évidentiel? *Langue Française* 102: 24–40. doi:10.3406/lfr.1994.5712
- Dendale, Patrick. 2000. *Devoir* épistémique à l'indicatif et au conditionnel: Inférence ou prédiction? In *Actes du XXIIe Congrès International de Linguistique et de Philologie Romanes, Bruxelles, 23–29 juillet 1998*, Annick Englebert, Michel Pierrard, Laurence Rosier & Dan van Raemdonck (eds), 159–169. Tübingen: Niemeyer.
- Erelt, Mati. 2001. Some notes on the grammaticalization of the verb *pidama* in Estonian. In *Estonian Typological Studies V*, Mati Erelt (ed.), 6–25. Tartu: Tartu University Press.
- Erelt, Mati & Metslang, Helle. 2009. Some notes on proximative and avertive in Estonian. *Linguistica Uralica* XLV 3: 178–191. doi:10.3176/lu.2009.3.02
- Gide, André. 1915. *If it die...*, trans. by Dorothy Bussy. London: Secker and Warburg.
- Gimpel, Jean. 1980. *Les bâtisseurs de cathédrales*. Paris: Seuil.
- Habicht, Külli. 2001. Vanhan kirjaviron modaalirakenteista. In *Congressus Nonus Internationalis Fenno-Ugristarum 7.–13.8.2000 Pars IV Dissertationes sectionum: Linguistica I*, Tõnu Seilenthal (ed.), 268–276. Tartu: Eesti Fennougristide Komitee.
- Habicht, Külli & Prillop, Külli. 2016. Polüfunktsionaalne *pidama*-verb XVII sajandi kirjakeeles (The multifunctional verb *pidama* in the 17th-century written Estonian). *Emakeele Seltsi aastaraamat* 62: 38–61.
- Kilgi, Annika. 2010. *Tõlkekeele dünaamika piibli esmaeestinduse käigus: Verbi morfosüntaksi areng ja lõplik toimetamisfaas* (The dynamics of language in the first translation of the Bible into Estonian). PhD dissertation, University of Tallinn.
- Kronning, Hans. 1996. *Modalité, cognition et polysémie: Sémantique du verbe modal 'devoir'*. Uppsala & Stockholm: Acta Universitatis Upsaliensis & Almqvist & Wiksell International.
- Kronning, Hans. 2001. Pour une tripartition des emplois du modal *devoir*. *Cahiers Chronos* 8: 67–84.
- Kross, Jaan. 2003. *The Czar's Madman*, trans. by Anselm Hollo. London: The Harvill Press.
- Kuteva, Tania. 2000. TAM-auxiliation, and the avertive category in Northeast Europe. In *Grammaticalisation aréale et sémantique cognitive: Les langues fenniques et sames, Actes du Colloque International du C.N.R.S. tenu les 9 et 10 avril 1999 en Sorbonne*, Jocelyne Fernandez-Vest (ed.), 27–41. Tallinn: Eesti Keele Sihtasutus.
- Le Matin*, daily paper.
- Lévi-Strauss, Claude. 1961. *Tristes Tropiques*, trans. by John Russell. London: Hutchinson & Co.
- Martin du Gard, Roger. 1922–1923. *Les Thibault I: Le cahier gris – Le pénitencier – La belle saison*. Paris: Folio (Gallimard).
- Metslang, Helle. 2002. -ki/-gi ka ja nende soome kaimud (-ki/-gi, ka and their Finnish relatives). *Lähivertailuja* 12: 57–81.
- Metslang, Helle. 1996. The development of the futures in the Finno-Ugric languages. In *Estonian: Typological Studies, I*, Mati Erelt (eds), 123–144. Tartu: Tartu University Press.
- Metslang, Helle. 1994. Eesti ja soome – futuurumita keeled? (Estonian and Finnish – Languages without future?). *Keel ja Kirjandus* 9–10: 534–616.
- Nilsson-Ehle, Hans. 1943–1944. Le conditionnel 'futur du passé' et la périphrase devait + infinitif. *Studia Neophilologica* 16: 50–88. doi:10.1080/00393274308586938

- Picoche, Jacqueline. 1993. *Didactique du vocabulaire français*. Paris: Nathan.
- Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York NY: Macmillan.
- Sartre, Jean-Paul. 1964. *Nausea*, trans. by Lloyd Alexander. New York NY: New Direction.
- Tammsaare, Anton Hansen. 2014. *Truth and Justice. Andres and Pearu*, trans. by Inna Feldbach and Alan Peter Trei. Stockholm: Haute Culture Books/Kreell AB.
- Treikelder, Anu. 2016. Eesti modaalverbi *pidama* episteemilistest kasutustest võrdluses prantsuse verbiga *devoir* (The epistemic use of the Estonian modal *pidama* in comparison with the French modal *devoir*). *Eesti ja soome-ugri keeleteaduse ajakiri – Journal of Estonian and Finno-Ugric Linguistics* 7(2): 159–186.
- Treikelder, Anu. 2015. La modalité épistémique dans une perspective contrastive: Les verbes modaux *devoir* français et *pidama* estonien. In *Actes du XIXème Congrès des Romanistes Scandinaves: Le XIXème Congrès des Romanistes Scandinaves, Reykjavik, 12.–15.Aug.2014*, Sigrún Ástríður Eiríksdóttir (ed.), 1–17. <<http://conference.hi.is/rom14/rom-lectures/>>
- van der Auwera, Johan & Plungian, Vladimir. 1998. Modality's semantic map. *Linguistic Typology* 2: 79–124. doi:10.1515/lity.1998.2.1.79
- Vetters, Carl. 2004. Les verbes modaux *pouvoir* et *devoir* en français. *Revue Belge de Philologie et d'Histoire* 82(3): 657–671. doi:10.3406/rbph.2004.4851
- Vetters, Carl & Barbet, Cécile. 2006. Les emplois temporels des verbes modaux en français: Le cas de *devoir*. *Cahiers de Praxématique* 47: 191–214.

The competition between the present conditional and the prospective imperfect in French over the centuries: First results

Jacques Bres, Sascha Diwersy and Giancarlo Luxardo
Praxiling, UMR 5267, Université Paul-Valéry Montpellier 3

While a significant number of works are known about the competition over the centuries between the synthetic future and the periphrastic prospective present, the parallel competition between the present conditional and the periphrastic construction with the prospective imperfect has received less attention. We focus here on the latter competition, limiting our study of the two forms to their temporal uses in discourse. We describe the uses for each form according to their order of appearance: imminence, subjective ulteriority and objective ulteriority. The two forms compete with each other in the expression of both subjective and objective ulteriority. We give an assessment of the diachronic evolution of the ratio present conditional / prospective imperfect in the expression of subjective ulteriority.

Keywords: prospective, subjective ulteriority, objective ulteriority, grammaticalization

1. Introduction

Extensive work has been carried out in the area of French grammar about the competition over the centuries between the synthetic future (*partira*, ‘leave-FUT-3SG’) and the periphrastic construction with the prospective present (*va partir*, ‘go-PRS-3SG leave-INF’) (e.g. Damourette & Pichon 1970 [1911–1936]; Fleischman 1982; Poplack & Turpin 1999; Poplack & Dion 2009; Veters & Lière 2009; Lindschouw 2011; Abouda & Skrovec 2015). The main subjects of these analyses are (i) the diachronic evolution of the grammatical forms, i.e. the cycle of future tenses (analytical form → synthetic form → analytical form) and (ii) the illustration of the grammaticalization process of a movement verb into a temporal auxiliary.

Much less attention has been paid to the parallel competition between the present conditional (*partirait*, ‘leave-COND-3SG’) and the periphrastic construction with the prospective imperfect (*allait partir*, ‘go-IMP-3SG leave-INF’), certainly because these forms occur less frequently.¹ Indeed, whereas different labels have been given to the construction *va partir* (e.g. *futur proche*, *futur périphrastique* denominations, *futur immédiat*), nothing similar had occurred for *allait partir* ‘go-IMP-3SG leave-INF’. Fleischman (1982), for instance, devoted a number of pages to the distinction between *partira* ‘leave-FUT-3SG’ and *va partir* ‘go-PRS-3SG leave-INF’, but none to the distinction between *partirait* ‘leave-COND-3SG’ and *allait partir* ‘go-IMP-3SG leave-INF’.

This chapter focuses on the relationship between the present conditional and the prospective imperfect in their temporal uses. We first provide an explicit account of the uses in discourse of each form according to their diachronic appearance: (a) subjective anteriority (10th century) and objective anteriority (19th century) for the conditional; (b) imminence (15th century), subjective anteriority (16th century) and objective anteriority (late 18th century) for the prospective imperfect. The two forms compete with each other in the expression of both subjective and objective anteriority. Choosing to only address the question of diachronic evolution of the ratio present conditional / prospective imperfect in the expression of subjective anteriority, we describe our corpus and the methodology we used (i.e. text retrieval and hierarchical clustering) for the experiment. Finally, we present and discuss the results of the study, including quantitative data and charts.

Preliminary comments should be made about the choice of the terms prospective present (*présent prospectif*: *va partir*) and prospective imperfect (*imparfait prospectif*: *allait partir*). This choice is based on the fact that the current denomination *periphrastic future* appears problematic for two reasons: first, it cannot be applied to *allait partir* as the auxiliary is in the imperfect; second, for both forms (present and imperfect), the denomination is relevant only for one meaning effect (admittedly frequent), i.e. anteriority, while the periphrasis also produces imminence as another meaning (cf. *infra* 2.2). We will use the term *prospective*, proposed by Benveniste (1959/1966: 245) or Comrie (1976: 64), which allows us to group the various meanings produced by the periphrasis.

1. To the best of our knowledge, only Sarrazin and Azzopardi (2012) have so far addressed this issue (contrastively in French and Spanish), in oral contemporary corpora.

2. Uses in discourse of the present conditional and the prospective imperfect

Although both the present conditional and the prospective imperfect are likely to have either aspecto-temporal or modal uses (Bres, Azzopardi & Sarrazin 2012), we will only focus on the former.

2.1 The present conditional

The present conditional may have two types of temporal uses in discourse, subjective ulteriority and objective ulteriority, which are diachronically very distant:²

Table 1. Conditional

| | | | |
|-------------|---------------------------|---|--------------------------|
| Conditional | subjective ULTERIORITY | → | objective ULTERIORITY |
| Stages | (a) (10th) | | (b) (19th) |

2.1.1 Stage (a) – subjective ulteriority

Already in early texts from the 10th century, the conditional is used to actualize an event as ulterior with respect to a past utterance. This past utterance may be explicit – indirect discourse as exemplified in (1) – or implicit – free indirect discourse as illustrated in (2).

- (1) *cil jurerent sus sainz lealment que il iroient* ('go-COND-3PL') *par les destroiz de Marroth.* (Villehardouin, *La Conquête de Constantinople*, 1208)
'they swore on the gospel saints that they would go loyally through the straits of Morocco.'
- (2) *Il s'allongea, les yeux au plafond. Il remonterait* ('go back up-COND-3SG') *là-haut à la nuit tombante.* (J. Carrière, *L'Épervier de Maheux*, 1972)
'He laid down, with his eyes to the ceiling. He would go back up there at dusk.'

Because ulteriority is assessed from an utterance, the process may then be realized or not: it falls within the scope of branching time (Kripke 1963; Rescher & Urquhart 1971; Gardies 1975; Martin 1981). As the past is irrevocable, it has a linear structure, whereas the future has a branching structure because it falls under what is possible. The Crusaders in (1) either went through the straits of Morocco or they did not; similarly, the character in (2) either went back up or he did not.

2. We borrow the adjectives *subjective* (\approx non-factive) and *objective* (\approx factive) from Nilsson-Ehle (1943).

2.1.2 Stage (b) – objective ulteriority

From the 19th century onwards, due to a decline in subjective mediation, the conditional may have indicated that an event is ulterior to another one in the objective unilinearity of the past (Damourette & Pichon 1970 [1911–1936]; Nilsson-Ehle 1943; Bres 2012).³ This use of the conditional is not frequent. In this case, it may be glossed with a simple past as in (3) and (4):

- (3) *La morale devait ensuite se trouver de plus en plus affectée par le discrédit croissant qu'allait nécessairement subir une théologie qui, désormais rétrograde, deviendrait* ('become-COND-3SG') *enfin radicalement antipathique à la raison moderne.* (A. Comte, *Discours sur l'esprit positif*, 1844)
 'Morality then became more and more affected by the growing discredit that would necessarily undergo a theology, which, henceforth retrograde, would finally become radically disagreeable to modern reason.'
- (4) *Soulagée de se voir marcher assez droit, Gloire préfèrerait* ('prefer-COND-3SG') *rentrer à pied plutôt que de prendre un taxi.* (J. Echenoz, *Les Grandes blondes*, 1995)
 'Relieved to be seen walking straight enough, Glory would prefer to return on foot rather than to take a taxi.'

In these two previous uses, the conditional is in competition with the prospective imperfect, and therefore it may be replaced by it (*remonterait / allait remonter; préfèrerait / allait préférer*).

2.2 The prospective imperfect

Aller is a prototypical instance of the grammaticalization of a spatial lexical unit into a grammatical unit indicating tense (e.g. Fleischman 1982; Heine & Kuteva 2002; De Mulder 2002; Bres & Labeau 2013a) according to the “verb-to-affix cline”, thus summarized by Hopper & Traugott (1993: 108): “full verb > auxiliary > clitic > affix”. In the present case, it is the first step (auxiliary). The movement verb *aller*, stage (a), is grammaticalized into an auxiliary of imminence, stage (b); then into an auxiliary of subjective ulteriority, stage (c); and finally, into objective ulteriority, stage (d), according to a development path that may be represented as follows:

3. From subjective constructions to objective constructions, the variation is modal, but it is limited to the scope of the aspecto-temporal value of ulteriority.

Table 2. Prospective imperfect

| | | | | | | | |
|--------|----------|---|------------|---|-------------|---|-----------------|
| Aller | MOVEMENT | → | IMMINENCE | → | subjective | → | objective |
| | | | | | ULTERIORITY | | ULTERIORITY |
| Stages | (a) | | (b) (15th) | | (c) (16th) | | (d) (late 18th) |

Two other development paths of the grammaticalization of *aller* should be pointed out: first, the narrative use (Gougenheim 1929; Bres & Barceló 2007), found as early as the 13th century as exemplified in (5), and second, the *allure extraordinaire* (since the 15th) (Damourette & Pichon 1970 [1911–1936]; Bres & Labeau 2013b) as instantiated in (6):

- (5) *Sur ces propos, feirent leur accord, et, en regardant le lieu le plus propre pour faire ceste belle oeuvre, elle vat dire* ('go-PRS-3SG dire-INF') *qu'elle n'en sçavoit point de meilleure ne plus loing de tout soupson, que une petite maison qui estoit dedans le parc, où il y avoit chambre et lict tout à propos.*

(M. de Navarre, *L'Heptaméron* 1559)

'About this subject, in agreement, and while searching for an appropriate place to do this beautiful work, she told him that she didn't know a better or further one from suspicion, than a small house in the park where there was a suitable room and bed.'

In this example of narrative use, *vat dire* (go-PRS-3SG say-INF) competes with the *passé simple* (preterit). This construction, attested between the 13th and 17th centuries, is grammaticalized as a periphrastic preterit only in Catalan and a few Occitan dialects.

- (6) *Elle oblya tellement son honneur et conscience, qu'elle alla aymer* ('go-PRET-3SG love-INF') *un jeune homme.* (M. de Navarre, *L'Heptaméron*, 1559)
'She forgot her honor and conscience to such an extent that she loved a young man.'

The *allure extraordinaire*, which is constructed with all tenses, has spanned over the centuries and is today quite alive.

2.2.1 Stage (a) – movement

Aller as a verb means the displacement of an animate agent towards a destination to be reached at the end of this movement, therefore at a later stage than this very movement (*Corinne allait* ('go-IMP-3SG') à Sète, 'Corinne was going to Sète'). This destination may be a process in the infinitive, which is realized at the end of this displacement, as in (7). This use, which is found in Vulgar Latin, is attested at all times.⁴

4. E.g. in the *Vulgate* by Jerome (382): *Veniunt adorare eum*, 'ils viennent l'adorer'.

- (7) *Et Tronc lez mist en prison, et lez alloit* ('go-IMP-3SG') *visiter moult songneusement*
(anonyme, *Ysaÿe le Triste*, 1400)
'Tronc put them in prison and went to visit them with care.'

2.2.2 Stage (b) – imminence

The movement creates a spatial gap – the process is reached at the end of this gap – which implies a temporal gap. By metonymy, one gets from the spatial gap to the temporal gap, to indicate the stage immediately preceding a process, namely, its imminence (as early as the 15th century), as in (8) and (9):⁵

- (8) *Je m'en voys à la ville; car je sçay bien que je trouveray tout prest à monter à cheval et mes gens aussi; car le cappitaine de Crathor les alloit tous faire habiller* ('go-IMP-3SG dress-INF'), *quant je suis sailly icy avec vous.*

(Jean de Bueil, *Le Jouvencel*, 1461)

'I'm leaving for the city; for I know that I would find my men ready to mount their horses. The captain Crator was about to dress them when I left with you.'

- (9) – *Il a des qualitez qui me le font vanter.*
– *Il ne les auroit pas s'il n'alloit heriter* ('go-IMP-3SG inherit-INF').
(P. du Ryer, *Les Vendanges de Suresne*, 1636)

'– He has qualities that make me praise him.

– He wouldn't have them if he wasn't about to inherit.'

Allait + *inf.* represents the pre-processual phase, and may be glossed by "*était sur le point de*". Therefore "*s'il n'alloit hériter*" ≈ "*s'il n'était sur le point d'hériter*", 'he wasn't about to inherit'.

The imminence may be avertive (Henry 1952; Kuteva 1998), in that the right co-text means that the process is not realized, as in (10):

- (10) *Elle m'allait traïter* ('go-IMP-3SG treat-INF') *en auteur du forfait, / Mais ma fuite a rendu sa fureur sans effet.* (P. Corneille, *La Veuve ou le Traître trahi*, 1634)
'She was about to treat me as the perpetrator of the crime. But my escape rendered her fury useless.'

5. According to Detges (1999), followed by De Mulder (2008) (and in agreement with Bybee *et al.* 1994, according to whom the grammaticalization of a form as a future goes through a development step, in which an *intention* is expressed), an *intentionality step* appeared as soon as the 13th century, that is, before the imminence step, by virtue of an inference starting from the displacement meaning: the agent moves *with the aim* of doing this or that. We have to admit that we are not convinced by the examples proposed by Detges and De Mulder to validate the hypothesis of an autonomous intentionality step: intentionality is assumed as early as the first step, and we have not found any occurrences for which intentionality does not go along with displacement.

2.2.3 Stage (c) – subjective ulteriority

From the representation of the pre-processual phase, one gets into the representation of ulteriority of the processual phase. And this is realized with an important enunciative modification: in order to produce the meaning of subjective ulteriority, the process must be in indirect speech, as shown in (11):

- (11) *Il est parti soudain après votre amour sue, /A trouvé tout aisé, m'en a promis l'issue, / Qu'il allait y pourvoir* ('go-IMP-3SG provide-INF') *et que vers la minuit /Vous fussiez toute prête à déloger sans bruit.*

(P. Corneille, *L'Illusion comique*, 1639)

'He left suddenly as soon as he knew that you loved him, he promised me that he was going to find a happy exit, and that around midnight you would be ready to leave silently.'

In (11), the prospective imperfect is in a complement clause and the ulteriority is thus assessed from the utterance (*promettre*) of an agent set in the past. The process will then be achieved or it will not; it falls into the scope of branching time.

2.2.4 Stage (d) – objective ulteriority

Starting from the end of the 18th century, the prospective imperfect in an independent clause may, as a sign of the fading of the utterance of an agent, locate an event as subsequent in the linearity of the past.

- (12) *Je ne m'étais jamais crue si près de la mort; hélas! Elle allait pourtant s'offrir* ('go-IMP-3SG happen-INF') *à moi sous un aspect encore plus réel. Roland revient, il me sort du cercueil.* (Sade, *Justine ou les Malheurs de la vertu*, 1791)

'I had never thought to be so close to death; alas! It was going to happen to me under an even more real appearance. Roland comes back, he gets me out of the coffin.'

In this case, the prospective imperfect may be glossed by a simple past (*passé simple*) or by a present perfect (*passé composé*):

- (13) *Je ne m'étais jamais crue si près de la mort; hélas! Elle s'offrit / s'est offerte* ('happen-RET-3SG'/'be-PRS-3SG happen-PTCP-PST') *pourtant à moi sous un aspect encore plus réel.*

'I'd never thought to be so close to death; alas! It happened to me under an even more real appearance.'

The prospective imperfect actualizes an event which has taken place. It is fully inscribed in the objective unilinear past.

With regard to this development path, it is worth mentioning that: (a) the new meanings do not eliminate the former meanings, but add up to them; (b) the

transition may be seamless: for instance, some occurrences may perfectly be relevant both to imminence and subjective ulteriority.

The competition between the prospective imperfect and the present conditional occurs in examples in stages (c) and (d): cases that do not represent the process in its pre-processual phase, as in stage (b), but in a processual phase referred to as ulterior, either subjectively (stage (c)) as in (11) or objectively (stage (d)) as in (12).

In order to follow diachronically the competition of the two forms, we now focus on subjective ulteriority, previously illustrated for the present conditional by Examples (1) and (2), and for the prospective imperfect by (11). As a consequence, since its use for both forms lies within the context of finite complement clauses and is introduced by the conjunction *que*, these content clauses can easily be retrieved in large amounts in a text corpus (compared to other uses).

3. An experiment with a multi-genre diachronic corpus of annotated texts

Our study is based on the large body of texts gathered within the framework of the *PRESTO* project (ANR-DFG program). This project produced a corpus of 545 text samples (monographs, collections, etc.) from the 16th to the 20th centuries, belonging to various discursive genres (narrative, poetry, theatre, treatise, correspondence, etc.). A part of these samples have been selected from the well-known archives *Frantext*, *Épistémon*, *ARTFL*, *CÉPM* (Corpus Électronique de la Première Modernité) and *Gallica* (for details on the processing of the *PRESTO* corpus, cf. Diwersy et al. 2015).

The entire corpus is categorized and lemmatized by means of a part-of-speech tagger (*TreeTagger*), which applies a linguistic annotation model designed and developed as part of the *PRESTO* Project. Table 3 provides an outline of the corpus we used.

In order to retrieve the clauses relevant to our study, we first ran a set of automated queries against the corpus (based on *CQP* – Corpus Query Processor, cf. Evert & Hardie 2011), targeting sequences composed of the conjunction *que* ('that') followed by a verbal form in the present conditional or prospective imperfect. We then randomly selected 50% of the sequences matching the query, keeping the relative proportions per century. Then, the selected sequences were searched for instances meeting the requirement of subjective ulteriority in a finite complement clause. As a result, we were able to build a concordance comprising 3,067 lines, which was submitted to statistical processing and is presented in the following section.

Table 3. Text samples comprising the PRESTO corpus

| Sample | Period | Genres | Texts | Words |
|------------|--------------|--------------------------------------------------------------------|-------|------------|
| PRESTO_16s | 16th century | narrative, poetry, theatre, treatise | 50 | 3,048,984 |
| PRESTO_17s | 17th century | narrative, poetry, theatre, treatise, correspondence | 106 | 7,042,386 |
| PRESTO_18s | 18th century | narrative, poetry, theatre, treatise, correspondence, encyclopedia | 128 | 7,762,787 |
| PRESTO_19s | 19th century | narrative, poetry, theatre, treatise, correspondence, press | 150 | 8,592,846 |
| PRESTO_20s | 20th century | narrative, poetry, theatre, treatise, press | 111 | 8,692,270 |
| Total | | | 545 | 35,139,273 |

4. Evolution of the competition between conditional and prospective

Table 4 shows the relative proportions per century of the present conditional and the prospective imperfect with respect to the occurrences observed in our concordance.

Table 4. Distribution of the two forms per century

| Verbal form | 16th | 17th | 18th | 19th | 20th | Total |
|----------------------|--------|--------|--------|--------|--------|--------|
| ALLAIT + INF | 3 | 22 | 58 | 54 | 158 | 295 |
| Conditional | 210 | 727 | 819 | 438 | 578 | 2772 |
| Total occurrences | 213 | 749 | 877 | 492 | 736 | 3067 |
| ALLAIT + INF (ratio) | 1.41% | 2.94% | 6.61% | 10.98% | 21.47% | 9.62% |
| Conditional (ratio) | 98.59% | 97.06% | 93.34% | 89.02% | 78.53% | 90.38% |

In the following paragraphs, we focus our observations on the prospective imperfect form, since it is the one that starts to compete with the conditional and the calculations related to the two forms are complementary. The previous figures reveal a steady increase in its proportion from the 16th up to the 20th century.

In order to examine the changes occurring in each century more finely, we re-processed the results shown in Table 4 and analyzed them as a series of observations based on 25-year intervals. Such chronological series may be analyzed according to a procedure called Variability-based Neighbor Clustering (VNC), introduced by Gries and Hilpert (2008, 2012) and already implemented for the study of diachronic

corpora. Figure 1 shows the VNC dendrogram for our data on the prospective imperfect, overlaid with a line plot representing the chronological evolution of its observed proportion in each of the 25-year intervals.⁶

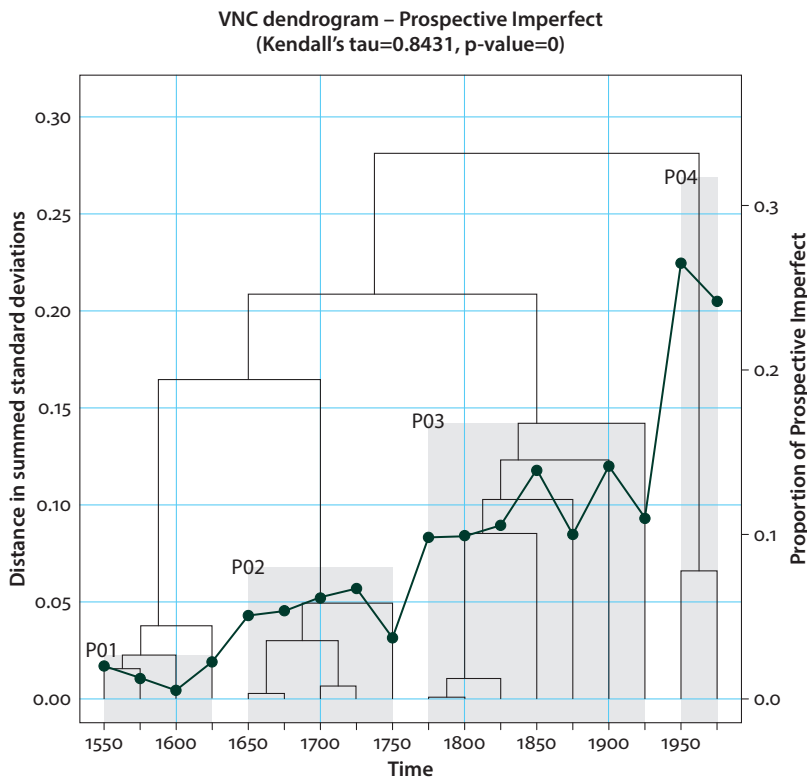


Figure 1. Variability-based neighbor clustering

A close examination of this hierarchical clustering leads to the highlighting of four adjacent areas (displayed in gray in the figure), representing four successive stages in the evolution of the prospective imperfect (see Table 5). The dendrogram has to be analyzed from the top:

- a first split appears on the year 1950, indicating the main tipping point (isolating stage P04 on the right),
- in the cluster on the left, the year 1775 marks a second split (revealing stage P03),
- the last split is around the year 1650 (between P01 and P02).

6. Figure 1 includes in its main title the value yielded by Kendall's Tau rank correlation coefficient, which indicates a significant upward trend throughout the whole period of investigation. The use of Kendall's Tau for analyzing global trends in diachronic data has been suggested by Hilpert and Gries (2009: 389–390).

If Figure 1 is read following the line of chronological evolution, two additional observations may be drawn from this figure: (a) throughout period P03 (and in particular between 1850 and 1950), the proportion increase of the prospective imperfect is characterized by some fluctuation; (b) the transition from P03 to P04 (second half of the 20th century) marks a neat quantitative shift in the progression of the prospective imperfect.

Table 5. Four stages in the evolution of the prospective imperfect

| Period | Cumulative proportion (by period) | |
|--------|-----------------------------------|--------|
| P01 | 1500–1649 | 1.49% |
| P02 | 1650–1774 | 5.56% |
| P03 | 1775–1949 | 11.00% |
| P04 | 1950–1999 | 25.05% |

Having established these trends of periodization, let us now turn to the linguistic features that might be involved in the selection of the prospective imperfect or the present conditional when used to express subjective ulteriority. We choose as a starting point for further investigation the semantic category of the word (verb, noun or adjective) that takes the subordinate clause as its complement. Close inspection of the retrieved concordance shows that we are dealing with two major classes: (a) words expressing the mediation of perception and thought (class P, as illustrated in (14)); and (b) words expressing the mediation of communication (class C, as illustrated in (15)).

- (14) *À l'angle de la petite allée qui conduit à la maison, il s'arrêta. On parlait dans la cuisine. Il sentit ('feel-PRET-3SG') que le souffle allait lui manquer. Ses jambes s'étaient remises à trembler. Il avança pourtant, l'oreille tendue.*

(B. Clavel, *Les Fruits de l'hiver*, 1968)

'In the angle of the small alley that leads to the house, he stopped. Somebody talked in the kitchen. He felt like the breath was going to leave him. His legs had started to tremble again. Yet he moved, listening carefully.'

- (15) *Après avoir bien disputé, on se sépara sans convention & sans accommodement: sur cela on ordonna ('order-PRET-3SG') que les actes de la conférence seroient supprimés, & qu'on garderoit le silence sur les matières contestées.*

(D. Diderot & J. d'Alembert (s.l.d.), *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, Tome 7, 1757)

'After having a discussion, they separated without agreement: on this, it was ordered that the conference proceedings be erased and that silent be kept on the contested materials.'

Testing the association between these two classes and the prospective imperfect in each of the diachronic stages, which we have identified, yields the following results

if the log-likelihood ratio is chosen as statistical test⁷ (in Table 6, the number of asterisks indicates different levels of significance, viz. in decreasing order ***: highly significant, **: very significant, *: significant). Positive values indicate attraction, whereas negative ones indicate repulsion.⁸

Table 6. Semantic categories in the four stages

| | C (communication) | P (perception and thought) |
|-------------------------------------------|-------------------|----------------------------|
| PROSPECTIVE IMPERFECT (P01: 1500–1649) | –15.72*** | 15.72*** |
| PROSPECTIVE IMPERFECT (P02: 1650–1774) | –16.8*** | 16.8*** |
| PROSPECTIVE IMPERFECT (P03: 1775–1949) | –7.93** | 7.93** |
| PROSPECTIVE IMPERFECT (P04: 1950–1999) | –3.95* | 3.95* |

These figures show that throughout all the stages there is a significant attraction of the prospective imperfect to the class of words expressing the mediation of perception and thought. In the last stage, however, this attraction is maintained in a barely significant way (lower values indicate less attraction).

In other words, there is evidence that in the earlier stages of its evolution, the prospective imperfect is co-textually more dependent on the semantic category of the words that govern the complement clause in which it appears; whereas in the most recent stage (which corresponds more or less to contemporary French), it tends toward more autonomy, which might indicate a certain kind of generalization of its subjective use. The questions remain then as to: (a) why the prospective imperfect is more likely to be found in complement clauses that represent reported perception and thought than in clauses that represent reported communication; and (b) whether a more fine-grained classification would reveal particular associations between the prospective imperfect and some subclasses (such as terms denoting perception in contrast to terms denoting cognition or emotion).

Besides testing the effects of a more fine-grained semantic classification, this analysis should be pursued with the inclusion of additional variables, but also with the observation of more than one variable at a time (multivariate analysis), which may provide new findings on the uses of these competitive forms. For now, we

7. The log-likelihood ratio is used here as a statistical test to estimate the correlation between the two variables (semantic category and time).

8. As we are dealing with a configuration of two variables with two levels each, the calculated scores are exactly the opposite in terms of attraction and repulsion.

simply make note of the fact that, since the beginning, the ratio of the prospective imperfect has been constantly growing with respect to the present conditional, and that this evolution goes hand in hand with the balancing of the distribution that holds between the two major semantic classes of the words governing the complement clause in which the prospective imperfect appears.

5. Conclusion

We elucidated the temporal uses in discourse of the present conditional and the prospective imperfect according to their diachronic appearance and their competition in the expression of ulteriority. First, we outlined in detail the specific configurations underpinning this competition, and then we chose to primarily address the question of the diachronic evolution of the relationship between the present conditional and the prospective imperfect, in the expression of subjective ulteriority within the scope of finite complement clauses introduced by the conjunction *que* (*content clauses*).

The data we extracted from a large corpus of text samples ranging from the 16th to the 20th century suggest that the proportion of *prospective imperfect* forms is steadily increasing over time and that this upward trend involves four successive stages. With respect to these four diachronic periods (ranging from 1500 to 1649, from 1650 to 1774, from 1775 to 1949 and from 1950 to 1999 respectively), we finally considered the influence of another variable, i. e. the semantic category of the introductory term, as it is likely to impact the competition between *present conditional* and *prospective imperfect*. We showed that this variable had some significance, at least in the early stages of the competition.

We intend to test possible correlations with other variables, such as discursive genre or the sociolinguistic dimension. So far, we have not been able to find evidence to support these hypotheses – the distribution of discursive genres found in the corpus studied here is related to the period of time.

Future steps of investigation may include several questions as follows. First, whereas both forms mean ulteriority, they are morphologically different: a synthetic form for the present conditional (which stems from the grammaticalization as an affix of *avoir* in the imperfect) and a periphrastic form for the prospective imperfect (which stems from a less advanced grammaticalization stage of the verb *aller* as an auxiliary). Does this morphological difference translate into a semantic difference when it comes to uses in discourse? In the corpus, is there a distribution of occurrences between the two that can be explained by a difference of temporal or enunciative representation of the process? Because of its derivation from the meaning of *imminence*, was the prospective imperfect initially (i.e. in the 16th and

17th centuries) used to signify imminent ulteriority? If so, has it lost the instruction of imminence over the centuries, in order to more widely signify the ulteriority in the past and clearly compete with the conditional present? As the prospective imperfect means the ulteriority of a process originating at the time of a reported enunciation act, can the difference at the enunciative level be explained along the lines suggested by Fleischman (1983: 190), i.e. the present prospective “express[es] the speaker’s subjective view at the moment of utterance”.

Second, is the competition in the expression of objective ulteriority similar to the competition in the expression of subjective ulteriority?

Third, does the competition affecting the temporal uses of the conditional also affect some of its modal uses?

Fourth, is the competition between present conditional and prospective imperfect similar to the competition between synthetic future and prospective present? Are both competitions based on the same semantic difference between the synthetic form and the periphrasis?

Fifth, to what extent does a new narrative construction developed in the 20th century (Bres & Labeau 2013c) and exemplified by (16) and (17), revive the old construction (5) or rather derive from the construction of ulteriority, both in the present and in the imperfect?

- (16) *Teddy Pendergrass est remarqué par Harold Melvin, leader du quintette vocal The Blue Notes. Il rejoint alors la formation, qui va enchaîner* (‘go-PRS-3SG go ON-INF’) *une succession de tubes. En 1976, Teddy Pendergrass décide de mener une carrière solo et quitte les Blue Notes. Il varégulièrement occuper* (‘go-PRS-3SG occupy-INF’) *les premières places des meilleures ventes de disques aux USA.*

(*Le Monde*, Obituaire de T. Pendergrass, 27.1.2010)

‘Teddy Pendergrass is noticed by Harold Melvin, leader of the vocal quintet The Blue Notes. He joins the group, who would go on to have a series of hits. In 1976, Teddy Pendergrass decides to start a solo career and leaves the Blue Notes. He would go on to regularly occupy the top of the best-selling discs in the USA.’

- (17) *Le ballon frisait le montant d’un Buffon qui paraissait battu (9e). Le défenseur de l’Inter, peu à la fête jusqu’à présent, allait néanmoins se racheter* (‘go-IMP-3SG redeem-INF’).

(*Le Figaro*, 09/07/2006)

‘The ball grazed the goalpost of Buffon who appeared defeated (9th). The defender of Inter, not really at the party up until now, would go on to nevertheless redeem himself’.

Finally, what happens to the competition between present conditional and prospective imperfect in other Romance languages, which have grammaticalized ‘go’ as a temporal auxiliary (e.g. Spanish or Portuguese)?

Corpus

- Anonyme, *Ysaÿe le Triste*. 1400.
 Jean de Bueil, *Le Jouvencel*. 1461.
 Jean Carrière, *L'Épervier de Maheux*. 1972.
 Bernard Clavel. *Les fruits de l'hiver*. 2002.
 Auguste Comte, *Discours sur l'esprit positif*. 1844.
 Pierre Corneille, *La Veuve ou le Traître trahi*. 1634.
 Pierre Corneille, *L'illusion comique*. 1639.
 Denis Diderot & Jean d'Alembert (s.l.d.), *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, Tome 7. 1757.
 Jean Echenoz, *Les Grandes blondes*. 1995.
 Marguerite de Navarre. *L'Heptaméron*. 1559.
 Pierre du Ryer, *Les Vendanges de Suresne*. 1636.
 Marquis de Sade, *Justine ou les Malheurs de la vertu*. 1791.
 Geoffroy de Villehardouin, *La Conquête de Constantinople*. 1208.
Le Figaro, 09/07/2006.
Le Monde, Obituaire de T. Pendergrass, 27.1.2010.

References

- Abouda, Lotfi & Skrovec, Marie. 2015. Du rapport entre formes synthétique et analytique du futur. Étude de la variable modale dans un corpus oral micro-diachronique. *Revue de Sémiotique et Pragmatique* 38: 35–57.
- Benveniste, Émile. 1959/1966. Les relations de temps dans le verbe français. *Bulletin de la Société de Linguistique LIV*, Fasc. 1: 237–257. Reprint In *Problèmes de linguistique générale*. Paris: Gallimard.
- Bres, Jacques. 2012. Conditionnel et ultériorité dans le PASSÉ: De la *subjectivité* à l'*objectivité*. *Actes du 3^{ème} Congrès Mondial de Linguistique Française (CMLF)*, Lyon: 1719–1730. doi:10.1051/shsconf/20120100037
- Bres, Jacques, Azzopardi, Sophie & Sarrazin, Sophie. 2012. Le conditionnel en français: Énonciation, ultériorité dans le passé et valeurs modales. *Faits de Langues* 40: 37–43.
- Bres, Jacques & Barceló, Gerard Joan. 2007. La grammaticalisation de la forme *itive* comme *prospectif* dans les langues romanes. In *Combat pour les langues du monde/Fighting for the world's languages: Hommage à Claude Hagège*, Jocelyne Fernandez-Vest (ed.), 91–104. Paris: L'Harmattan.
- Bres, Jacques & Labeau, Emmanuelle. 2013a. *Aller et venir*: Des verbes de déplacement aux auxiliaires aspectuels-temporels-modaux. *Langue française* 179: 13–28. doi:10.3917/lf.179.0013
- Bres, Jacques & Labeau, Emmanuelle. 2013b. *Allez donc sortir des sentiers battus!* La production de l'effet de sens *extraordinaire* par *aller* et *venir*. *Journal of French Language Studies* 23(2): 151–177. doi:10.1017/S0959269512000154
- Bres, Jacques & Labeau, Emmanuelle. 2013c. The narrative construction *va + infinitive* in contemporary French: A linguistic phoenix rising from its medieval ashes? *Diachronica* 30(3): 295–322. doi:10.1075/dia.30.3.01bre

- Bybee, Joan L., Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago IL: The University of Chicago Press.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: CUP.
- Damourette, Jacques & Pichon, Édouard. 1970 [1911–1936]. *Des mots à la pensée: Essai de grammaire de la langue français*, Tome V. Paris: d'Artrey.
- Detges, Ulrich. 1999. Wie entsteht Grammatik? Kognitive und pragmatische Determinanten der Grammatikalisierung von Tempus Markern. In *Reanalyse und Grammatikalisierung in den romanischen Sprachen*, Jürgen Lang & Ingrid Neumann-Holzschuh (eds), 31–52. Tübingen: Max Niemeyer.
- De Mulder, Walter. 2002. Grammaticalisation, métaphore et métonymie: Le cas d'aller. *Verbum* 24(3): 229–246.
- De Mulder, Walter. 2008. Grammaticalisation, métonymie et pertinence. *Actes du 1^{er} Congrès Mondial de Linguistique Française*: 359–365. <<https://doi.org/10.1051/cmlf08323>>
- Diwersy, Sascha, Falaise, Achille, Lay, Marie-Hélène & Souvay, Gilles. 2015. Traitements pour l'analyse du français préclassique. *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles*: 565–571. Caen: Association pour le Traitement Automatique des Langues. <http://www.atala.org/taln_archives/TALN/TALN-2015/taln-2015-court-037.pdf>
- Evert, Stefan & Hardie, Andrew. 2011. Twenty-first century corpus workbench: updating a query architecture for the new millennium. *Proceedings of the Corpus Linguistics 2011 Conference*, University of Birmingham, UK. <<https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf>>
- Fleischman, Suzanne. 1982. *The Future in Thought and Language*. Cambridge: CUP.
- Fleischman, Suzanne. 1983. From pragmatics to grammar: Diachronic reflections on complex pasts and futures in Romance. *Lingua* 60: 183–214. doi:10.1016/0024-3841(83)90074-8
- Gougenheim, Georges. 1929. *Etude sur les périphrases verbales du français*. Paris: Nizet.
- Gardies, Jean-Louis. 1975. *La logique du temps*. Paris: Presses Universitaires de France.
- Gries, Stefan T. & Hilpert, Martin. 2008. The identification of stages in diachronic data: Variability-based neighbor clustering. *Corpora* 3(1): 59–81. doi:10.3366/E1749503208000075
- Gries, Stefan T. & Hilpert, Martin. 2012. Variability-based neighbor clustering: A bottom-up approach to periodization in historical linguistics. In *The Oxford Handbook of the History of English*, Terttu Nevalainen & Elizabeth Closs Traugott (eds), 134–144. Oxford: OUP.
- Heine, Bernd & Kuteva, Tania. 2002. *World Lexicon of Grammaticalization*. Cambridge: CUP. doi:10.1017/CBO9780511613463
- Henry, Albert. 1952. Du subjonctif d'imminence contrecarrée à un passage du Tristan de Bérουλ. *Romania* 73: 392–407. doi:10.3406/roma.1952.333010.3406/roma.1952.3330
- Hilpert, Martin & Gries, Stefan T. 2009. Assessing frequency changes in multi-stage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing* 24(4): 385–401. doi:10.1093/lc/fqno12
- Hopper, Paul J. & Traugott, Elizabeth. 1993/2003. *Grammaticalization*. Cambridge: CUP.
- Kripke, Saul A. 1963. Semantical considerations on modal logic. *Acta Phil. Fennica* 16: 83–94.
- Kuteva, Tania. 1998. On identifying an evasive gram: Action narrowly averted. *Studies in Language* 22(1): 113–160. doi:10.1075/sl.22.1.05kut
- Lindschouw, Jan. 2011. L'évolution du système du futur du moyen français au français moderne: La réorganisation comme un cas de régrammation. *Revue de Linguistique Romane* 75 (297): 51–98.

- Martin, Robert. 1981. Le futur linguistique: Temps linéaire ou temps ramifié? (A propos du futur et du conditionnel en français). *Langages* 64: 81–92. doi:10.3406/lgqe.1981.1886
- Nilsson-Ehle, Hans. 1943. Le conditionnel ‘futur du passé’ et la périphrase *devait* + infinitif. *Studia Neophilologica* 16: 50–88. doi:10.1080/00393274308586938
- Poplack, Shana & Turpin, Danielle. 1999. Does the *futur* have a future in (Canadian) French? *Probus* 11: 133–164. doi:10.1515/prbs.1999.11.1.133
- Poplack, Shana & Dion, Nathalie. 2009. Prescription vs. praxis: The evolution of future temporal reference in French. *Language* 85(3): 557–587. doi:10.1353/lan.0.0149
- Rescher, Nicholas & Urquhart, Alasdair. 1971. *Temporal Logic*. Vienna: Springer. doi:10.1007/978-3-7091-7664-1
- Sarrazin, Sophie & Azzopardi, Sophie. 2012. L’alternance du conditionnel et de la périphrase *itive* à l’imparfait dans des corpus oraux espagnols et français. *Studia UBB Philologia*. LVII(3): 57–68.
- Vetters, Carl & Lière, Audrey. 2009. Quand une périphrase devient temps verbal: Le cas d’*aller* + infinitif”. *Faits de Langue* 33: 27–36.

Evidentiality and the TAM systems in English and Spanish

A cognitive and cross-linguistic perspective

Juana I. Marín Arrese

Universidad Complutense de Madrid

This chapter examines and proposes plausible explanations for attested extensions of tense-aspect-modality (TAM) systems to inferential and reportative evidentiality based on the contrastive case study – British English (BrE) vs. Peninsular Spanish (PSp) – of some TAM markers realizing evidential functions: epistemic modals ‘must’, ‘must have’, and ‘*deber (de)*’, ‘*deber (de) haber*’; future perfect verbal forms ‘will have’, ‘*habrá/n*’; and conditional perfect verbal forms ‘would have’, ‘*habría/n*’. It is argued that the parameters playing a crucial role in facilitating these extensions include immediacy in relation to the ground and reality (cf. Langacker 2017) for inferential values, while irrealis (cf. Givon 1989) and frame-shifting to a counter point of view (Chilton 2014) are relevant for reportative evidentiality.

Keywords: evidentiality, inferential, reportative, modality, TAM systems

1. Introduction

This chapter examines the relation of TAM systems to inferential and reportative evidentiality. From a Cognitive Grammar account, it is assumed that temporal and modal meanings are best conceived of as instantiations of a common epistemic schema, where inference is a basic component of both evidentiality and the TAM systems (cf. Brisard 2013; Langacker 2017).

Extensions from the domain of epistemic modality to inferential evidential values have been posited for epistemic necessity modals, English *must* and Spanish *deber (de)*, as well as for the modal/future maker *will* (van der Auwera & Plungian 1998; Squartini 2004; Boye 2012, *inter alia*). Boye (2012: 137ff) has argued that these synchronically polyfunctional modal expressions represent extensions from the subspace or notional region of epistemic support (epistemic modality) to that

of indirect epistemic justification (inferential evidentiality), within the semantic map of epistemicity.

Inferential values are also found in association with the future simple and perfect, and the conditional simple, while the reportative value is characteristically found with the conditional perfect and also with the conditional simple in Spanish (Squartini 2001; Brisard 2013; Jaszczolt 2013; Vatrican 2015; Marín-Arrese 2017b).¹ It is argued that certain parameters play a crucial role in facilitating these extensions. Immediacy in relation to the ground² (immediate vs. non-immediate), reality (real vs. projected) (cf. Langacker 2017) are mainly involved in the extension to inferential evidentiality. It is also posited that the semantic extension of the conditional perfect to the indirect-reportative subdomain may be doubly motivated by conceptual distance from the ground, both in terms of non-immediacy and of the feature irrealis (Givón 1989), and counterfactuality or frame-shifting to a counter point of view (Chilton 2014). Non-immediacy facilitates the shift from personal to non-personal involvement, and the irrealis and counterfactuality features contribute to enhance the speaker's diminished responsibility concerning evidentiary justification, and the conception of an alternative mental space.

The chapter presents results of a contrastive case study, British English (BrE) vs. Peninsular Spanish (PSp), on some of these tense-aspect-modal markers realizing evidential functions. The aim is to identify, describe and provide a plausible explanation for the main attested extensions of the following TAM markers to epistemic and indirect-inferential evidentiality, and to indirect-reportative values of evidentiality in the two languages: (a) inferential evidential values of epistemic modals, simple and perfect forms: 'must' + INF, 'must have' + PP, and 'deber (de)' + INF, 'deber (de) haber' + PP; (b) epistemic and inferential evidential values of future perfect verbal forms: 'will have' + PP, 'habrá/n' + PP; (c) reportative evidential values of conditional perfect verbal forms: 'would have' + PP, 'habría/n' + PP.

The chapter also aims to explore the degree to which these evidential values are associated with particular discourse domains (written journalistic vs. unscripted conversation). The data consist of naturally occurring examples, randomly selected

1. The case of perspective-switching with the past imperfective to express free indirect speech is beyond the scope of this paper (cf. Saussure & Morency 2012). The simple future and conditional tenses with epistemic/inferential senses are also beyond the scope of this paper, for reasons explained in the Methodology section. The simple conditional in Spanish may also extend to the reportative domain, typically in a reportative context: *Según una teoría evolucionista, Quintano sería <IRE> el eslabón perdido del mudéjar de Serrano, ... (CREA) (According to an evolutionist theory, Quintano would be <IRE> the missing link of the Mudejar style in Serrano, ...).*

2. Langacker (2017: 52) notes that the ground comprises "the interlocutors, the speech event, and their location in time, space, and reality".

from spoken and written corpora in the two languages (BNC-Baby, CORLEC-UAM, CESJD-JMA).

This chapter is organized as follows. Section 2 discusses the main parameters involved in the extensions of TAM expressions as markers of inferential and reportative evidentiality. The case study is described in Section 3. The results and discussion are presented in Section 4. The final section is devoted to the conclusions.

2. Evidentiality and the TAM systems in English and Spanish

2.1 Dimensions and parameters of evidentiality

Evidentials allude to or specify the sources of knowledge and modes of access to the information whereby the speaker feels entitled to make an assertion or claim (Anderson 1986; Chafe 1986; Aikhenvald 2004). In the words of Anderson (1986: 277), “evidentials are used (a) to specify factual claims and (b) to indicate the justification available to THE PERSON MAKING THE CLAIM”. There is relative consensus in the literature regarding the various categories of evidential systems; a basic distinction is made between direct and indirect evidentiality, the latter comprising indirect inferential, and indirect reportative evidentiality (Diewald & Smirnova 2010).

Various dimensions and parameters have been proposed as criteria for the classification of evidential values or functions (cf. Plungian 2001; Squartini 2008). Plungian (2001) distinguishes between sources and forms of access to the information (direct vs. indirect evidence) and speaker’s involvement (personal vs. mediated evidence). Squartini (2008: 918) argues for the need to distinguish between “the mode of knowing or type of evidence as opposed to the source of evidence”. The source of evidence refers to the locus of information, as internal or external to the speaker/writer. The modes of knowing refer to the process whereby the information is accessed, either directly through sensory perception, or indirectly through inferences or reports. Tournadre & LaPolla (2014: 241) propose a definition of evidential marking as “the representation of source and access to information according to the speaker’s perspective and strategy”, where the notion of source is restricted to “primarily a verbal source of information (reported information)”, and access refers specifically to “non-verbal access to information (sensory, inferential, etc., including the sensory access to verbal source)”.

Drawing on these notions, we may conclude that the dimensions and parameters which appear to be crucial in distinguishing the values and subcategories of evidentiality are the following: (a) ‘source’ of the evidence, originating either in SELF or in OTHER (+ personal/-personal); (b) modes of knowing: direct access

to the evidence as opposed to indirect means, such as inference or reports; (c) type of evidence or experiential domain of the evidence: perceptual, conceptual, communicative (cf. Marín-Arrese 2013; Cornillie, Marín-Arrese & Wiemer 2015).

Langacker (2017: 20) argues that in striving for mental or epistemic control to cope with the world, we build up a conception of reality, which is constantly modified by access to knowledge through some basic sources of information: internal experience, sensory perception, higher-level cognition, as well as other conceptualizers. Bearing in mind that egocentricity is at the very centre of our mental universe, it follows that epistemic assessment would involve centrality with respect to a series of dimensions: time (speaker/ conceptualizer's *now*), source of information (internal experience, perception), and social interaction (access to other conceptualizers' minds). Different facets of these dimensions are manifested in the grounding system, as Langacker (2017: 21) observes: "A tense-modal system emphasizes time and reality, whereas an evidential system highlights the source and reliability of information". Centrality with respect to the ground in tense-modality systems is characterized by the basic distinctions "immediate vs. distant" in time, and "real vs. projected". Langacker (2017: 52) notes that "in evidential systems, a basic distinction is direct vs. indirect knowledge, the former including perception, and the latter both inference (mediation by higher-level cognition) and report (information from other individuals)". We argue here that these dimensions related to egocentricity and centrality, immediacy (immediate vs. distance from the ground), reality (real vs. projected), and the dimension *realis* vs. *irrealis*, will be parameters crucially involved in facilitating the extensions of these TAM markers to indirect-inferential and indirect-reportative values of evidentiality.

2.2 TAM systems in European languages: Epistemicity and inferentiality

TAM forms are attested for epistemic and evidential uses in a number of European languages, including Bulgarian, English, French, German, Italian, Portuguese and Spanish (Palmer 2001; Dendale 2010; Squartini 2001, 2004; Kronning 2015; Celle 2005, 2008; Laca 2006; Smirnova 2013; Oliveira 2015; Vatrican 2015, *inter alia*). As Cruschina and Remberger (2008: 98) note,

[T]he Romance languages express evidentiality through various modal and temporal forms, following a pattern that is typologically characteristic of the languages of Western Europe, where the development of evidentials from modal morphemes is a typical feature. [...] The conditional, the future, the imperfect, and periphrases with modal verbs (e.g. *dovere/devoir/deber* + infinitive), are traditionally assumed to take on evidentiality meanings in specific contexts. [...]

In English, both modal ‘must’ and ‘will’ have been attributed epistemic and evidential values. This notion is already present in Coates (1983: 41) who notes that “[e]pistemic ‘must’ conveys the speaker’s confidence in the truth of what he is saying, based on a deduction from facts known to him (which may or may not be specified)”. Similarly, Palmer (2001: 25) observes that one of the two main contrasts in the English modal system is that which “distinguishes between an inference from experience or general knowledge, i.e., between Deductive (‘must’) and Assumptive (‘will’)”:

- (1) [Knock at the door] That’ll be the postman. (Palmer 1986: 62)
- (2) a. It’s nine o’clock – John **will** be in his office now.
 b. Yes, the lights are on, so he **must** be there. (Palmer 2001: 28)

Within the cognitive linguistics paradigm, epistemic modality is analyzed in terms of the notion of force manifested at the mental level (cf. Sweetser 1982; Talmy 1988). This notion of potency at the mental level is conceived as the conceptualizer’s mental activity in his or her assessment of the “evolutionary momentum of reality” (Langacker 1991: 274), and interpreted as the force of evidence, which leads the speaker to the most likely conclusion. In the words of Langacker (1991: 274), these modal senses “might well engender the conception of the speaker being carried by the force of evidence along a deductive path”.

Langacker (1991: 246) notes that “the zero and distal forms of the modals respectively locate the designated process in *immediate* and *non-immediate irreal-ity*”. Modals ‘must’ and ‘will’ situate the process “at varying distances from the speaker’s position at immediate known reality”. The basic distinction between their present-time epistemic meanings, according to Langacker (1991: 280), derives from the conditions of confirmation of knowledge: “‘must’ conveys immediacy” and “(on the strictest interpretation) indicates that confirmation is regarded as virtually inevitable”, whereas “‘will’ suggests a certain epistemic distance” and “implies that confirmation requires a non-negligible expansion of present knowledge (so that new information might alter the prediction)”, as illustrated in the following examples from Langacker (*ibid*):

- (3) a. Since you saw it with your own eyes, it {must/*will} exist.
 b. They {will/*must} quite possibly be there by now.

In this vein, Squartini (2008: 922, 924) observes that a characteristic feature of epistemic modal ‘must’, in common with evidentials, is that it very often involves circumstantial inferential access to knowledge, and sometimes generic inference, in addition to its epistemic modal meaning of degree of certainty:

- (4) a. [Pointing to two people] They look like identical twins. They must/*will be brother and sister.
 b. [It is ten o'clock. The doorbell rings] That must/will be the postman.

The type of reasoning process in (4a) draws on a circumstantial inference³ (Anderson 1986: 284), which involves an inferential step from some given puzzling facts in the form of an external sensory evidence to some hypothesis that may explain it (cf. Givon 1989). In (4b), where both modals ‘must’ and ‘will’ are acceptable, there is an external sensory trigger (i.e. the door-bell rings) for the generic inferential process, which is based on previous personal experience of the speaker/writer or general world knowledge (i.e. the time at which the postman usually comes). In terms of logical reasoning, one might posit that the specific inference in (4a), based on results or facts, involves an “abductive”⁴ (Peirce 1955) process, while one might claim that (4b) exemplifies deduction, in that the premises lead to the logical conclusion. As such, one could argue that an abductive inferential process appears to be involved in core perceptual-based inferential evidentiality, whereas the type of deductive reasoning found in generic inferences would be more associated with epistemic necessity. Though modal ‘must’ shares the same functional-conceptual substance domain with the other modals (cf. Boye & Harder 2009), a comparable distinction³ between modals ‘must’ and ‘will’ lies in their acceptability in contexts (4a) and (4b): whereas ‘must’ is found with both specific and generic inferences, associated respectively with evidential and epistemic necessity, ‘will’ is restricted to generic inferences. This complementary distribution is similarly found in Spanish for the modal *deber* (*de*) and the future tense form, though the latter appears to be marginally acceptable for specific inferences as well. There appears to be certain variation in judgments of grammaticality in the literature on Romance languages, where acceptability seems to differ across languages in the case of contexts lacking external sensory-based evidence (cf. Tasmowski 2001; Dendale 1994; Squartini 2008):

3. Anderson (1986: 284) makes a distinction between “strong logical inferential ‘must have V-ed’”, “circumstantial inferential ‘seems to have V-ed’”, and a space corresponding to “weak inferential” located between the space of circumstantial inference and the space of “conjecture ‘maybe’”.

4. Peirce (1955: 151) provides the following general syllogism for abduction:

“The surprising fact, C, is observed;
 But if A were true, C would be a matter of course,
 Hence, there is reason to suspect that A is true.”

Peirce (1955) posits three distinct methods of reasoning, or stages in scientific research: inductive, deductive and abductive. Deductive reasoning is that which necessarily follows from general or universal premises, from generally accepted statements or facts. Abductive reasoning involves an inferential process of arriving at the best possible hypothesis or conclusion on the basis of knowledge at your disposal.

- (5) a. [Pointing to two people] *Parecen gemelos idénticos. Deben ser/??Serán hermano y hermana.*
 ‘They look like identical twins. They must be/?? They will be brother and sister’.
- b. [It is ten o’clock. The doorbell rings] *Debe ser/Será el cartero.*
 ‘That must be/That will be the postman’.

Inferential processes also comprise conjectures or ‘weak inferences’ (cf. Anderson 1986: 285), as in (6), that is, inferences occurring in contexts lacking any evidence, neither based on external sensory evidence nor triggered as part of general world knowledge, which is typically found with modals of low certainty, *may* or the distal form *might* (cf. Squartini 2008: 924). Again we observe certain differences between English and Spanish in the degree of acceptability of the use of modals and future forms in this context:

- (6) a. [The doorbell rings] Strange! I was not expecting anybody. It *must/*will/may/might be my neighbour.
- b. [The doorbell rings] ¡*Qué raro! No esperaba a nadie. ??Debe ser/??Será/Puede que sea/Podría ser mi vecino.*

Inferential processes, albeit distinct ones, are present in the use of all epistemic modals, and this points to obvious connections between epistemic modality and evidentiality. The three inferential processes, as Squartini (2008: 925) observes,

[...] are differentiated along a parameter that is connected to the balance between the speaker’s involvement as opposed to the import of external evidence. While in one of the two poles (circumstantial inferences), the speaker’s own reasoning is heavily supplemented by external sensory evidence, in the opposite pole, all external evidence is missing, the speaker being solely responsible for the reasoning process. In between, the intermediate area of generic inferences can be found with a balanced proportion of the speaker’s own reasoning and external information deriving from general world knowledge.

This inferential gradient posited by Squartini (2008), and previously by Anderson (1986), is distinct from a similar scale in terms of strength of knowledge, the epistemic scale, though there are obvious connections between the two. As Squartini (2008: 926) notes, “an inferential process based on external sensory evidence corresponds to a stronger epistemic commitment, while the speaker’s conjectures, lacking any sensory support, are intrinsically weaker if evaluated on an epistemic scale (Givón 1982)”. Thus, circumstantial inferences – involved in indirect perceptual-based evidentiality – and generic or logical inferences are characteristically associated with epistemic certainty and greater confidence, while weak inferentials or conjectures are associated with lesser confidence and epistemic possibility.

Modal *will* seems to lie within the intermediate territory of generic inferences and conjectures, as may be surmised from Example (3b) ('They {will/*must} quite possibly be there by now'), and Examples (4)–(6). The interplay of circumstantial and generic or logical inferential values is clearly exemplified in Example (7) (perceptual and conceptual evidence underlined) with modal 'must'; in (8) with *deber* (*de*) the perceptual-based evidence ('because there's one of those little horses that move ... Little lead soldiers') is also explicitly invoked:

- (7) <hit text = "KB7" n = "755"> She was screaming and you could hear her crying. Then you hear, she pounded up the bloody stairs and he was after her and I reckon well the only room with a lock As you say is the bathroom, is the bathroom, so she <kw>**must**</kw> <IIE>⁵ have been in there cos he's hammering on the door and bashing it and kicking it. Bloody hell of a row. Or whether she opened it or whether she was in the front bedroom with something behind the door I dunno but </hit> (BNC-B)
- (8) <H4> ¿Aquí venden gorros o qué? <H3> No. [[Deben]] <IIE> ser juguete. Yo... porque hay un caballito de esos que se mueven... Soldaditos de plomo... <H4> Mira hay... ahí sí que hay un caballito que... (CORLEC, cedu020d.asc)
<H4> Do they sell hats here or what? <H3> No. They **must** be toys. I ... because there's one of those little horses that move ... Little lead soldiers ...

Another distinctive feature associated with the two scales, the inferential and the epistemic, involves the dimension of subjectivity and intersubjectivity. The epistemic scale is intrinsically subjective, and the reasoning process is internal with respect to the speaker. However, in the inferential gradient, we find a distinction between conjectures, which involve an internal source and subjective reasoning, and circumstantial inferences, based on external evidence, which might be potentially intersubjectively shared with others, and which motivate a more objective kind of reasoning. In the intermediate area of generic inferences, we find an overlap between subjective, internal experience and the more intersubjective world knowledge (cf. Squartini 2008). This feature of intersubjectivity, associated with 'manifest observable evidence' (Sanders & Spooren 1996), which is at the core of evidentiality, may be playing a crucial role in the preferential extension of 'must' over the other modals to the inferential evidential function. Modals in English, as grounding elements, are inherently subjective; they thus evoke the speaker as the implicit conceptualizer, and this feature appears to effectively block further extension of modal *must* to the domain of reportative evidentiality, as has also been

5. The annotation system for the tokens found in the corpora is the following:

- <IIE> Indirect-Inferential Evidentiality
- <IRE> Indirect Reportative Evidentiality

ascertained for Spanish *deber* ('be').⁶ In contrast, the evocation of a generalized or virtual conceptualizer, as is the case of core evidential resources, verbal forms ('seem', 'appear') and adverbials ('apparently'), may be a factor in facilitating the extension to reportative values (cf. Marín-Arrese 2017b).

Epistemic modality and evidentiality may thus be considered complementary dimensions of epistemic assessment and the striving for epistemic control (cf. Langacker 2013). As Langacker (2017: 20) observes,

[e]vidential systems and tense-modality systems are therefore neither sharply distinct nor mutually exclusive. They are better seen as overlapping and complementary strategies of clausal grounding, providing an epistemic assessment concerning the existential status of the profiled occurrence. Individually or jointly, they help fulfil the referential function of a clause by way of achieving epistemic control and inter-subjective alignment.

2.3 Temporality-aspectuality and epistemicity

Langacker (1991) posits that at the most schematic level of definition, tense markers indicate epistemic distance from speaker/conceptualizer's position; the distinction between present and past tenses may basically be conceived of as an epistemic distinction between immediate reality (IR) and non-immediate reality (NIR). Temporal concepts, Jaszczolt (2009: 140–141) argues, parallel a scale of epistemic modality reflecting degrees of "detachment from the certainty of now", from highest certainty (least detachment) to lowest certainty (most detachment). Similarly, Chilton (2013: 253) notes that across languages, it is relatively common to find uses of future tense forms which in certain contexts yield an epistemic, or "putative" cognitive effect, as a result of the "close cognitive connection between future time reference and the inherent cognitive (or even metaphysical) uncertainty of the future". However, Chilton cautions that this connection does not warrant conflating the conceptual dimensions of temporality and modality.

The epistemic future (EF), according to Saussure & Morancy (2012: 209) may be explained by positing an imaginary future perspective, or allocentric viewpoint, which allows for "a future verification of a possible state of affairs in the present". They make reference to the relevance-theoretic proposal by Sthioul (1998), who conceives of the representation of this future verification as "an imaginary allocentric point of consciousness in the future, from which the truth of the eventuality can be verified and therefore assessed as a fact" (Saussure & Morancy 2012: 210).

6. This implies no universality claim on my part. I am grateful to one of my anonymous reviewers for pointing out that "The modal verb 'ska' in Swedish (meaning both "must" and "will") shows a clear reportative value".

Drawing on Sthioul (1998) and Saussure & Morancy (2012), Chilton (2013) applies the frame-shift model of Deictic Space Theory (DST) to provide an explanation of how modal effects may arise in certain contexts. Chilton (2014: 164) argues for the reflection of time onto modality, for the mapping in deictic space of the t -axis (temporality) onto the m -axis (epistemic experience): “relative temporal distance is mapped onto relative epistemic distance, all in conceptual pre-linguistic space”. The future tense form ‘will’ in English typically denotes a confident prediction of some event in the future, so “it is plausible to see it as mapping reflectively onto a relatively ‘close’ region on the epistemic axis” (Chilton 2014: 165). For the epistemic or putative future, Chilton (2013: 256) posits a “meta-represented point of view at some $t_1 > t_0$ ”, which “does not have to be an unspecified allocentric other”; rather, this additional point of view is “that of a cognitive avatar of S , S' , and can still be regarded as ‘allocentric’ in that sense”. This explanation, as Chilton (2013: 256) notes, is also consistent with the contexts where “it is pragmatically manifest that the speaker herself will be in a position to verify, in the future, a current situation”. This is clearly the case in the canonical example for assumptive ‘will’ in (1), where a speaker might reason:⁷ ‘In some (near) future, when I open the door, in all probability I will be able to assert truthfully “It is the postman calling”’. This feature of potential verification may also be a crucial factor for the epistemic effect of the future in contexts involving some inferential calculation about a present eventuality.

In the case of the future perfect, Chilton (2013: 256) argues that the putative reading also involves “verification in the future relative to S ”, but in this case the location of the event is “in the past relative to S , i.e. at some $t_1 < t_0$ ”. The construction simultaneously represents reference to a past event relative to the ground (S) in the ‘have been’ component, and future reference with respect to the ground in the ‘will’ component.

(9) Henry will have visited/will have been visiting Calais.

It would thus appear that non-immediacy with respect to the ground and potential verification play a crucial role in facilitating the extension of the future and future perfect tense forms to epistemic values. The following are examples of the future perfect in English and the future and future perfect in Spanish with epistemic and evidential inferential values.

7. Saussure (2013: 63) spells out the prediction conveyed by (1) in the following terms: “In some (near) future, it will be possible to utter truthfully ‘It is the postman.’”

- (10) "... But the danger is that what starts off being benevolent can, and usually does, deteriorate into being malevolent." That is not a view that [[will have]] <IIE> been heard with any great pleasure in Downing Street. A judicial gauntlet has been thrown down. I wonder whether Mrs Blair would consider picking it up.
(CESJD-EOT)
- (11) <H2> Claro, pero... imagi<palabra cortada>... seguramente **habrá** <IIE> tantos varones porque [[habrá]] <IIE> **habido** algún <simultáneo> infanticidio de mujeres... <H1> ...<palabra cortada>ticidio. Claro. </simultáneo> <H2> ... porque tampoco es natural... que haya ve<(i)>ntiuno... <risas> <simultáneo> en proporción.
(CORLEC, ahum033a.asc)
- <H2> Of course, but ... I imagine there **will** probably be so many males because there [**will have**] **must have** been some infanticide of women ...)

These extensions are not random since they are attested in systems with grammatical evidentiality, as well as in systems lacking a fully grammaticalized system of evidentiality such as Spanish and other Romance languages. In discussing circumstantial-inference evidentials, Anderson (1986: 275) makes the point that these often arise "by historical change from certain kinds of perfect (those of result state or of current relevance)". The effect of inference based on a verifiable resultant state no doubt increases the potential of the extension of the future perfect to indirect inferential evidentiality.

With regard to the conditional, the if-clause, as Chilton (2014: 163) notes, may be considered an "inherently modalising" element, since it has scope over the apodosis and locates it at some distance from the ground in a space of irrealis. The distal forms of the modals (e.g. 'would', 'might') indicate greater epistemic distance from the ground; so modal 'would' "specifies a kind of non-immediacy with an imaginative component" (Langacker 2017: 29). Examples of the conditional perfect form with an epistemic reading found in both languages are the following; Example (13) is particularly interesting, with its interplay of conditional simple with present-time epistemic sense and the perfect with an epistemic value and resultative meaning:

- (12) ... it might not have been built but it er it's Victorian so it could have been built like in eighteen thirty soon fa certainly soon after eighteen thirty that <kw>**would have**</kw> <IIE> been built.
(BNC-B)
- (13) *El Partido Socialista Obrero Español **tendría** <IIE> entre veinte y veintiuno lo cual quiere decir que hubiera... <vacilación> que [[habría]] <IIE> sufrido un descenso... bastante notable. E Izquierda Unida, como hemos venido reflejando también, **subiría** <IIE> a seis o siete.*
(CORLEC, enot003g.asc)
- (The Socialist Party **would** have between twenty and twenty-one which means it would ... **would have** suffered a drop ... quite considerable. And United Left, as we have also been showing, **would** rise to six or seven)

As we have seen from the above discussion, the epistemic future seems to involve epistemic distancing from reality as well as potential future verification. The epistemic future perfect similarly involves distance from some past reference point relative to the ground, and potential future verification relative to that past reference point. The epistemic conditional and conditional perfect would involve an additional step in non-immediacy, involving the feature irrealis, which would result in further epistemic distance and a lesser degree of realness.

2.4 Temporality-aspectuality and reportativity

The function of reportative evidentiality is to signal that access to the information is through some source external to the speaker/conceptualizer. Reportativity represents an alternative source of indirect evidence, distinct from egocentricity, since it no longer involves speaker's assessment but rather mediation through another conceptualizer, an external perspective.

We may assume that the different types of conditional sentences correspond to scalar variations in non-immediacy and irrealis. Chilton (2014: 159) observes that conditional sentences have “the potential to express degrees of irrealis, ranging from the not-quite-certain about the near future to that which is counter to fact”. Counterfactual conditionals “could, then be defined as the *furthest* ‘possible world’ (‘mental space’ in cognitivist terminology) relative to the position of the speaker” (ibid: 162).

The semantic extension of the conditional perfect to the indirect-reportative subdomain may be doubly motivated by conceptual distancing from the ground, both in terms of non-immediacy or distance from known reality and also in terms of the feature irrealis, in the sense of locating events in the realm of the unreal, in hypothetical contexts or worlds. As Langacker (2017: 29) observes, in this case of non-immediate projection, “the profiled occurrence is projected from an imagined ground (G') within an alternate reality (R)'”.

Chilton (2014: 167) posits that epistemic distancing effects in conditional sentences may be modelled geometrically as mental spaces (cf. Fauconnier & Turner 2002), one representing the basic coordinate system, “S's deictic space R”, and the other second set of axes, or “‘new reality’ space”, which involves projecting “a copy R' to various points on S's *m*-axis”. Counterfactuality, as has been argued within cognitive semantics (Fauconnier & Turner 2002; Chilton 2014), involves the conceptualisation of two, incompatible and simultaneous, mental representations, an epistemically real representation and one counter to it. Conditionals evoke potential “‘alternative realities’”, whereas “in counterfactuals proper, the reality is *counter*, it is polarity-reversed” (Chilton 2014: 176–177). Counterfactual conditionals

presuppose the falsity of the expressed assertions in both the protasis and the apodosis, as in the following example from Chilton (2014: 173):

- (14) If John had gone to the party, he would have met Sarah.

However, this mental representation is countered by an alternative representation in an imaginary world, where the expressed assertions might be true. Chilton (2014: 173) proposes a reflection model⁸ of counterfactuals to explain this construction as “a configuration in which, from S’s point of view, John is at the party in R’ (*if John had been at the party*), but not at it in the presupposition (or better, in S’s base epistemic state)”. This would account for the simultaneity and contrariness of the two mental representations: the representation of ‘going to the party’ and ‘meeting Sarah’ is in the realis plane in R, and in the irrealis plane is R’. Thus, the speaker simultaneously builds a mental representation in a secondary imaginary (irrealis or counterfactual) space R’, where events and relations between entities run counter to those in the base space R (S’s real world).

This chapter argues that this notion of a ‘counter’ isometric world, as an irrealis space R’, may well have been the basis for the re-conceptualisation of this ‘counter’ world as S’s ‘mediated space’, R². The frame-shifting to a space of counter-reality, involving a shift from S’s perspective in R to the implicit perspective of some other S, S² in R², may be conceived as involving some potential ‘otherness’, the perspective of a cognitive ‘counter-avatar’ of S, S². The following example of the conditional perfect form with a reportative meaning may well derive its extended meaning from the re-analysis of the basic notions involved in counterfactuality: S²’s perspective and potential verification in an alternative ‘counter’ world, or secondary irrealis space R², is paralleled to the perspective of the covert source in mediated evidentiality.

- (15) *El ex jefe de guerra cristiano y ex ministro libanés Elie Hobeika, asesinado ayer [[habría]] <IRE> dejado cintas grabadas con revelaciones sobre la matanza de civiles palestinos en Sabra y Chatila y que aportaría en el juicio que se va a celebrar en Bruselas contra el ...* (CESJD-SNA)
 (The ex Christian war lord and Libanese ex Minister Elie Hobeika, murdered yesterday [[**would have**]] had **reportedly** left tape recordings with revelations concerning the massacre of Palestinian civilians in Sabra and Chatila ...)

8. Chilton (2014: 174) observes that this configuration is a “glide reflection: a reflection in the mirror perpendicular to *m* and passing through its limit point, combined with a translation of the reflected axes such that O’ coincides with the limit of *m* and the limit of *m*’ coincides with O. The reflection of the space R’ interlocks with the base space R.”

The notions posited here of the re-conceptualisation of a counter world as a ‘*mediated space*’, R^2 , and the perspective of a cognitive ‘*counter-avatar*’, S^2 , which parallels the shift from personal to non-personal involvement in mediated evidentiality, may well have paved the way for the extension of the conditional perfect to indirect reportative meaning. It is also assumed that this crucial distinction between hypothetical conditionals and counterfactuals may account for the preferred forms of extension: epistemic and inferential evidential extensions for the simple conditional form, and evidential reportative extensions for the conditional perfect.

3. The case study: Hypotheses, research objectives and methodology

The case study presents results of TAM markers (epistemic modals, simple and perfect forms, future perfect, and conditional perfect forms) realizing epistemic and inferential evidential, and reportative functions of evidentiality in the two languages:

- a. English: ‘*must*’ + INF, ‘*must have*’ + PP, ‘*will have*’ + PP, ‘*would have*’ + PP.
- b. Spanish: ‘*deber (de)*’ + INF, ‘*deber (de) haber*’ + PP, ‘*habrá/n*’ + PP, ‘*habría/n*’ + PP.

The study explores the presence of tokens of TAM expressions with epistemic or inferential evidential and reportative values attested in written journalistic discourse and in unscripted oral discourse.

3.1 Hypotheses and research objectives

The case study posits the following hypotheses:

H1: Extensions from the modal systems in English and Spanish to inferential evidential functions will be attested to a greater extent in modals invoking a strong inferential or conclusional force, and in the resultative perfect forms.

H2: Extensions from the tense-aspect systems in English and Spanish will differ in their potential for extension to the epistemic/inferential or to the reportative domain due to differences in the meaning potential of the linguistic resources: Non-immediacy and frame-shifting, characteristic of the future perfect, will favour epistemic/inferential values, whereas the features irrealis and counterfactuality, characteristic of the conditional perfect, will favour reportative values.

H3: There will be variation across discourse domains in the presence and degree of occurrence of evidential values expressed by TAM expressions, due to the distinct features of oral vs. written communicative events: a higher incidence of

multifunctionality in unscripted conversation is expected, since the oral language is the natural site for innovation and semantic extension.

In order to test the hypotheses, the following research objectives are defined:

- a. Identification, annotation and quantification of the selected expressions of tense-aspect-modality markers realizing inferential and reportative functions of evidentiality, as well as their non-evidential occurrences, in naturally occurring discourse;
- b. Comparison of the similarities or differences in the use of these resources across discourse domains (oral vs written discourse) and languages (English vs. Spanish).

3.2 Case study: Corpora

The data consist of naturally occurring examples selected from the following corpora:

- *Corpus of English and Spanish Journalistic Discourse* (CESJD-JMA), 2000–2010, a comparable corpus of journalistic texts (opinion columns, leading articles, and news-reports) in the two languages (English: 342,197 words; Spanish: 334,294 words).
- *BNC-Baby-Unscripted conversation* (English). (1,000,000 approx.)
- *Corpus Oral de Referencia de la Lengua Española Contemporánea* (CORLEC), Universidad Autónoma de Madrid. Mostly unscripted conversation (1,100,000 approx.)

An additional selection of data for the conditional perfect form in Spanish was subsequently taken from the CREA corpus, since only a limited number of tokens were obtained from the above corpora. The search was carried out in the subsections for Oral (Spain) and Newspapers (Spain):

- *Corpus de referencia del español actual* (CREA). Real Academia Española. <<http://www.rae.es>> (Newspaper texts dated 2000–2004: 8,474,325 words) (Oral: 3,214,296 words).

Since the Oral (Spain) subcorpus of CREA also includes the CORLEC corpus, the search included all the tokens minus the corresponding tokens from the CORLEC corpus.

3.3 Research design: Data collection and procedure

All the examples of tense-modality tokens present in the corpora were selected, analysed and manually tagged according to the target values specified above. A three-way distinction was established for the annotation system: T/AM (non-epistemic modal meanings and basic tense/aspect values); E/I (epistemic/inferential values); IRE (indirect reportative values). An electronic search using Monoconc was carried out to ensure that all the instances present in the texts had been identified. The data were submitted to further analysis for the quantitative results.

For the selection of the data the following criteria were applied:

- a. The choice of forms was restricted to the simple and perfect forms indicated above in order to facilitate the selection of all the instances present in the corpora chosen. Applying electronic searches for the simple conditional form in Spanish poses a problem, since the ending in *-ría* may appear in a multitude of words (a search in CORLEC yielded 3,125 matches, including nouns such as *cafetería*, *ingeniería*, etc.), which would make replicability with larger corpora very problematic.
- b. Only the present tense form (*debe*) and the perfect (*debe haber*) were included for the modal verb *deber* (*de*) in Spanish.
- c. Only 3SG/PL forms were included for Spanish TAM forms.
- d. Idiomatic expressions functioning as a complement-taking predicates (CTPs) were excluded: e.g. ‘I would have thought’.
- e. In the case of repetitions and instances of self-repair, only one token was included in the counts.

4. Results and discussion

Tables 1 and 2 below present results of the case study on epistemic/inferential and reportative evidential values of the tense-aspect-modality markers selected for the study. The T/AM column includes the non-epistemic uses of modals ‘must’ and *deber* (*de*), and in the case of tense-aspect forms it provides the basic tense/aspect values of the tokens found; the E/IIE column gives the epistemic/inferential uses, and the IRE column shows the indirect reportative values of the tokens.

4.1 Results: Epistemic modal > inferential evidentiality

Inferential evidential values were regularly found for epistemic necessity modals, English *must* and Spanish *deber* (*de*). In both languages we find a marked preference for these epistemic/inferential uses in oral unscripted conversation, as predicted in H3, whereas in journalistic discourse the tokens predominantly have non-epistemic necessity meanings (the number of non-epistemic tokens is almost ten times that of epistemic uses in English; in Spanish it is approximately forty-four times the number of tokens). This significant difference in frequency may be the result of discourse-domain and genre-related pragmatic and stylistic features of journalistic discourse, which often favours effective stance markers (cf. Marín-Arrese 2013).

Table 1. Modality markers in English and Spanish: Epistemic/Inferential and Reportative Evidentiality (raw numbers and ratio per thousand words)

| MODAL | CESJD 342,197 words | | | | BNC-BABY 1,000,000 approx. | | | |
|----------------------------|------------------------|-------|-----|-------|-------------------------------|-------|-----|-------|
| | T/AM | E/IIE | IRE | Total | T/AM | E/IIE | IRE | Total |
| ENGLISH | | | | | | | | |
| <i>must</i> | 287 | 29 | 0 | 316 | 204 | 279 | 0 | 483 |
| | 0.839 | 0.085 | | 0.923 | 0.204 | 0.279 | | 0.483 |
| <i>must have/of</i> | 0 | 15 | 0 | 15 | 0 | 182 | 0 | 182 |
| | | 0.044 | | 0.044 | | 0.182 | | 0.182 |
| Total | 287 | 44 | 0 | 331 | 204 | 461 | 0 | 665 |
| MODAL | CESJD 334,294 words | | | | CORLEC 1,100,000 approx. | | | |
| | T/AM | E/IIE | IRE | Total | T/AM | E/IIE | IRE | Total |
| SPANISH | | | | | | | | |
| <i>debe/n (de)</i> | 280 | 8 | 0 | 288 | 257 | 112 | 0 | 369 |
| | 0.838 | 0.024 | | 0.862 | 0.234 | 0.102 | | 0.335 |
| <i>debe/n (de) + haber</i> | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| | | | | | | 0.002 | | 0.002 |
| Total | 280 | 8 | 0 | 288 | 257 | 114 | 0 | 371 |

In English, epistemic/inferential values are for the most part expressed by modal ‘must’, and secondarily by the perfect form of the modal, ‘must have’. However, whereas with the simple form ‘must’ we find other non-epistemic modal meanings, the perfect form, ‘must have’, appears to function exclusively with an epistemic/inferential value, as predicted in H1. No doubt, the conclusional force of ‘must’ is increased when the reference to immediate known reality has an additional resultative meaning. In Spanish, epistemic/inferential values are also found for the modal verb *deber*, though their ratio of occurrence is substantially lower than that

of English *must*, and with only a marginal occurrence of the perfect form, which also appears to specialise for the epistemic/inferential function.

As expected with these grounding elements, in keeping with H2, no extensions to reportative values were found for the modal forms in English or Spanish. Elements such as modals implicitly ground the proposition in the contingent individual subjectivity of the speaker/conceptualizer (cf. Langacker 1991). This basic feature of modals in English, and presumably also in Spanish, appears to block any extension to reportative evidentiality, which is characterized by an indirect mediated access to the evidence originating in some non-specific external source.

4.2 Results: Future perfect and conditional perfect > epistemic/inferential values

As discussed above, the English tense-modal system also allows for inferential evidential values with the modal/future marker 'will', which typically denotes a confident prediction of some event in the future. The future perfect involves further distancing, both in terms of projection and temporal non-immediacy (cf. Langacker 2017). As we can see in Table 2, the future perfect in English, 'will have', shows very low ratios for the 'putative' reading in both discourse domains, and the ratio of usage is very similar to the tense-aspect values.

Table 2. TAM markers in English and Spanish: Evidential uses of future perfect and conditional perfect (raw numbers and ratio per thousand words)

| TAM | CESJD 342,197 words | | | | BNC-BABY 1,000,000 approx. | | | |
|-----------------|---------------------|-------|-------|-------|----------------------------|-------|-------|-------|
| | T/AM | E/IIE | IRE | Total | T/AM | E/IIE | IRE | Total |
| will have | 13 | 19 | 0 | 32 | 3 | 1 | 0 | 4 |
| | 0.038 | 0.055 | | 0.094 | 0.003 | 0.001 | | 0.004 |
| would have/of | 122 | 14 | 0 | 136 | 266 | 25 | 0 | 291 |
| | 0.357 | 0.041 | | 0.398 | 0.266 | 0.025 | | 0.291 |
| Total | 135 | 33 | 0 | 168 | 269 | 26 | 0 | 295 |
| TAM | CESJD 334,294 words | | | | CORLEC 1,100,000 approx. | | | |
| | T/AM | E/IIE | IRE | Total | T/AM | E/IIE | IRE | Total |
| <i>habrá/n</i> | 9 | 12 | 0 | 21 | 26 | 83 | 0 | 109 |
| | 0.027 | 0.036 | | 0.063 | 0.024 | 0.075 | | 0.99 |
| <i>habría/n</i> | 0 | 2 | 19 | 51 | 11 | 2 | 2 | 15 |
| | 0.089 | 0.006 | 0.057 | 0.153 | 0.010 | 0.002 | 0.002 | 0.014 |
| Total | 39 | 14 | 19 | 72 | 37 | 85 | 2 | 124 |

In Spanish, epistemic/inferential values of the future perfect, *habrá/n*, also show similar ratios to those of the tense-aspect usage in journalistic discourse, though it is considerably more frequent in the discourse domain of unscripted oral conversation, as predicted in H3.

The conditional perfect also realizes epistemic and inferential evidential functions in both languages. Results for English, ‘would have’, show only a marginal use for epistemic/inferential values in both discourse domains in comparison to its basic tense/aspect values in conditional sentences. In Spanish, the occurrence of the conditional perfect with an epistemic/inferential value is even more negligible.

As argued above, it would appear that non-immediacy to the ground, and potential future verification of an event located in the past with respect to the ground, play a role in facilitating the extension from future perfect to epistemic/inferential evidential values. No instances of extension to the reportative domain were found for the future perfect in either language, nor of the conditional perfect in English. The features non-immediacy and potential verification, which appear to be crucially involved in the epistemic/inferential evidential extensions of the future and future perfect, do not seem to warrant further extension to the reportative domain.

In English we have found no instances of extensions of TAM forms to the reportative domain, as results in Table 2 show. Reportative values in English appear to be restricted to lexical strategies (verbs, adverbials: ‘is reported to’, ‘allegedly’, ‘apparently’, ‘reportedly’, etc.). It is conceivable that extension to reportative values may somehow be blocked by the presence of modal auxiliaries (‘will’, ‘would’), which are grounding elements, in the future and conditional tense forms.

In Spanish, as predicted in H2, the future perfect, *habrá/n*, is restricted to epistemic/inferential values. In the case of the conditional perfect, *habría/n*, we find dual extensions to both the inferential and the reportative domains, with a considerably higher ratio in journalistic discourse, and equal ratios for inferential and reportative values in the oral discourse, no doubt due to the potential of conditionals to express “degrees of irreality” (Chilton 2014: 159).

4.3 Results: Conditional perfect > indirect reportative values

Reportative values of the conditional perfect in Spanish, *habría/n*, appear to be more entrenched in journalistic discourse, with only a marginal number of cases in unscripted oral discourse, as can be seen from the results for the CORLEC corpus in Table 3. These results are confirmed by the additional case study of the conditional perfect in the CREA corpus, where a substantial part of the oral corpus includes data from radio and TV programmes.

Table 3. TAM markers in Spanish: Evidential uses of conditional perfect (raw numbers, ratio per thousand words, and percentages)

| TAM | CESJD 334,294 words | | | | CORLEC 1,100,000 approx. | | | |
|-----------------|-------------------------------------------------|-------|-------|-------|------------------------------------|-------|-------|-------|
| | T/AM | E/IIE | IRE | Total | T/AM | E/IIE | IRE | Total |
| <i>habría/n</i> | 30 | 2 | 19 | 51 | 11 | 2 | 2 | 15 |
| | 0.089 | 0.006 | 0.057 | 0.153 | 0.010 | 0.002 | 0.002 | 0.014 |
| | 58.8% | 3.9% | 37.2% | 100% | 73.3% | 13.3% | 13.3% | 100% |
| | CREA (News, Spain) 8,474,325 words ⁹ | | | | CREA (Oral, Spain) 3,214,296 words | | | |
| | T/AM | E/IIE | IRE | Total | T/AM | E/IIE | IRE | Total |
| <i>habría/n</i> | 227 | 12 | 130 | 369 | 51 | 6 | 15 | 72 |
| | 0.027 | 0.001 | 0.015 | 0.043 | 0.016 | 0.002 | 0.005 | 0.022 |
| | 61.5% | 3.2% | 35.2% | 100% | 70.8% | 8.3% | 20.8% | 100% |

⁹ The information regarding number of words for CREA, Real Academia Española, can be found at: <<http://ntlle.rae.es/nomina/jsp/NominaFor.jsp>>

The results for the conditional perfect would run counter to H3. There appears to be a cline in the use of the conditional perfect with a reportative value, which goes from oral unscripted conversation to written journalistic discourse. It would seem that the conditional perfect with an indirect-reportative value is becoming firmly established in the context of journalistic discourse in Peninsular Spanish, in news reports on the radio or on TV, and has permeated and acquired a high degree of conventionalization in written journalistic discourse, as can be seen from the results of our corpora.

Results from the CREA corpus seem to confirm H2 in that the preferred extension for the conditional perfect in Spanish is to indirect-reportative values. Reportative evidentiality is characterized by an indirect mediated access to the evidence originating in some non-specific external source. As such the crucial feature which may be prompting this extension appears to be counterfactualty, since it involves frame-shifting to a space of counter-reality (Chilton 2014), which may have been re-conceptualised as that of the perspective of a cognitive ‘*counter-avатар*’ of S, S², and thence to an external perspective and mediated evidentiality.

9. The total number of words for the Press-Spain corpus (which includes both newspapers and magazines) is 39,596,727. In order to be able to recover all the instances of ‘*habría*’, the search was limited to the period 2000–2004, for Newspapers-Spain. The number of words for this period for Press-Spain is 8,474,325. It was not possible to recover the number of words for only newspapers, so the ratios are only tentative. No information is provided regarding the exact number of words in the Oral-Spain corpus, for which I made a personal inquiry. I have also given percentages of functions of tokens for *habría/n* in the three corpora used for Spanish, CORLEC, CESJD & CREA, for further comparison between them in Table 3.

The feature irrealis may also facilitate cases of extension of conditionals to the reportative domain in Spanish. Celle (2008: 24) observes that “[w]ith the conditional, the speaker disclaims responsibility for the propositional content which is claimed to emanate from a different assertive source and potentially discordant points of view coexist”. From a discourse-pragmatic perspective, the feature irrealis, presenting the situation as unrealized, contributes to distancing the speaker from the situation and diminishes speaker’s responsibility for the communicated information. As Givon (1989: 130) notes, “Under the scope of the IRR-assertion mode, information is weakly asserted [...]. The source of the information is thus largely irrelevant, since the speaker does not intend to defend the information too vigorously against challenge”.

The shift to some other point of view, from personal to non-personal, mediated evidentiality, allows the speaker to disclaim responsibility for the information. Weakly justified evidence, characteristic of reportative evidentiality, also shares the feature low or lack of subjective responsibility since the ‘displaced’ responsibility is attributed to an external voice (cf. Marín-Arrese 2017a). In heteroglossic discourse, these markers provide an excellent resource for the journalist to introduce the perspective of text external voices, without overtly attributing information to a specific secondary voice or quoted source. (cf. White 2012; Marín-Arrese 2017a).

5. Conclusions

This chapter has aimed to identify and describe some of the main extensions of tense-aspect-modality markers to express evidential values in English and Spanish, and to provide plausible explanations for those attested extensions in terms a set of dimensions and parameters. Crucial parameters potentially involved in facilitating these extensions have been argued to be potency and conclusional force, non-immediacy, irrealis and counterfactuality (cf. Chilton 2013, 2014; Langacker 2017).

The conclusional force associated with perceptual-based circumstantial inference appears to be central in the evidential readings of epistemic modal verbs. The additional resultative meaning of the perfect form ‘*must have*’ appears to increase the conclusional force, and renders the inferential evidential reading the favoured case. Since modals are inherently subjective, and evoke the speaker as the implicit conceptualizer, this feature appears to effectively block further extension of the modals to the domain of reportative evidentiality.

In the case of tense-aspect markers, non-immediacy, projected reality and the potential for future verification by frame-shifting to an ‘allocentric’ point of consciousness of a cognitive avatar of S, S’, would appear to be pave the way to epistemic/inferential uses (cf. Sthioul 1998; Saussure & Morancy 2012; Chilton 2013).

The feature irrealis (cf. Givon 1989), and the shift from S's perspective in R (reality space) to the implicit perspective of some other S, S' in R' (counter-reality), characteristic of counterfactual conditionals (Chilton 2014), may have motivated the re-conceptualization as the perspective of a cognitive 'counter-avatar' of S, S², in an alternative 'mediated space', R², thus paving the way for the extension of the conditional perfect to indirect reportative meaning.

The presence and distribution in discourse domains is very similar for both languages. Unscripted oral conversation seems to be the preferred site for epistemic/inferential values of modal verbs in both languages. In the case of the tense-aspect systems, epistemic/inferential values with the future perfect in English are practically restricted to written discourse; the conditional perfect, however, shows relatively similar ratios in oral and written discourse. Spanish clearly favours unscripted conversation for epistemic/inferential values with the future perfect, and journalistic discourse for the conditional perfect with reportative values. No definite conclusions, however, can be drawn from the data in relation to occurrence of the values in the discourse domains.

Acknowledgements

This study has received support from the project *Evidentiality: A Discourse-Pragmatic Study of English and other European Languages* (EVIDISPRAG), Ref. FFI2015-65474-P, funded by the Ministerio de Economía y Competitividad (Spanish Ministry of Economy and Competitiveness).

References

- Aikhenvald, Alexandra. 2004. *Evidentiality*. Oxford: OUP.
- Anderson, Lloyd B. 1986. Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In *Evidentiality: The Linguistic Coding of Epistemology*, Wallace Chafe & Johanna Nichols (eds), 273–312. Norwood NJ: Ablex.
- Boye, Kasper. 2012. *Epistemic Meaning: A Crosslinguistic and Functional-cognitive Study*. Berlin: Mouton de Gruyter. doi:10.1515/9783110219036
- Boye, Kasper & Harder, Peter. 2009. Evidentiality. Linguistic categories and grammaticalization. *Functions of Language* 16(1): 9–43. doi:10.1075/fol.16.1.03boy
- Brisard, Frank. 2013. An account of English tense and aspect in Cognitive Grammar. In *Time. Language, Cognition and Reality*, Kasia M. Jaszczolt & Louis de Saussure (eds), 210–235. Oxford: OUP. doi:10.1093/acprof:oso/9780199589876.003.0011
- Celle, Agnès. 2005. The French future tense and English *will* as markers of epistemic modality. *Languages in Contrast* 5(2): 181–218. doi:10.1075/lic.5.2.02cel
- Celle, Agnès. 2008. Tense, modality and commitment in modes of mixed enunciation. *Belgian Journal of Linguistics* 22(1):15–36.

- Chafe, Wallace. 1986. Evidentiality in English conversation and academic writing. In *Evidentiality: The Linguistic Coding of Epistemology*, Wallace Chafe & Johanna Nichols (eds), 261–272. Norwood NJ: Ablex.
- Chilton, Paul. 2013. Frames of reference and the linguistic conceptualization of time: Present and future. In *Time. Language, Cognition and Reality*, Kasia M. Jaszczolt & Louis de Saussure (eds), 236–258. Oxford: OUP. doi:10.1093/acprof:oso/9780199589876.003.0012
- Chilton, Paul. 2014. *Language, Space and Mind: The Conceptual Geometry of Linguistic Meaning*. Cambridge: CUP. doi:10.1017/CBO9780511845703
- Coates, Jennifer. 1983. *The Semantics of the Modal Auxiliaries*. London: Croom Helm.
- Cornillie, Bert, Juana I. Marin-Arrese & Wiemer, Bjorn. 2015. Evidentiality and the semantics-pragmatics interface. An introduction. *Belgium Journal of Linguistics* 29: 1–18.
- Cruschina, Silvio & Remberger, Eva-Maria. 2008. Hearsay and reported speech: Evidentiality in Romance. *Rivista di Grammatica Generativa* 33: 95–116.
- Dendale, Patrick. 1994. Devoir épistémique, marqueur modal ou évidentiel? *Langue Française* 102: 24–40. doi:10.3406/lfr.1994.5712
- Dendale, Patrick. 2010. *Il serait à Paris en ce moment. Serait-il à Paris? A propos de deux emplois épistémiques du conditionnel*. Grammaire, syntaxe, sémantique. In *Liens linguistiques. Etudes sur la combinatoire et la hiérarchie des composants*, Camino Alvarez Castro, Flor María Bango de la Campa & María Luisa Donaire Fernández (eds), 291–317. Bern: Peter Lang.
- Diewald, Gabriele & Smirnova. 2010. *Evidentiality in German. Linguistic Realization and Regularities in Grammaticalization*. Berlin: Mouton de Gruyter. doi:10.1515/9783110241037
- Fauconnier, Gilles & Turner, Mark. 2002. *The Way We Think*. New York NY: Perseus Books.
- Givón, Talmy. 1982. Evidentiality and epistemic space. *Studies in Language* 6(1): 23–49. doi:10.1075/sl.6.1.03giv
- Givón, Talmy. 1989. *Mind, Code, and Context: Essays in pragmatics*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Jaszczolt, Kasia M. 2009. *Representing Time: An Essay on Temporality as Modality*. Oxford: OUP.
- Jaszczolt, Kasia M. 2013. Temporality and epistemic commitment: An unresolved question. In *Time. Language, Cognition and Reality*, Kasia M. Jaszczolt & Louis de Saussure (eds), 193–209. Oxford: OUP. doi:10.1093/acprof:oso/9780199589876.003.0010
- Kronning, Hans. 2015. El condicional epistémico de atribución en francés, italiano y español: Aspectos diafásicos, diatópicos y diacrónicos. In *Les variations diasystématiques et leurs interdépendances dans les langues romanes*, Kirsten Jeppesen Kragh & Jan Lindschouw (eds), 507–521. Strasbourg: Société de linguistique romane/ÉLiPhi.
- Laca, Brenda. 2005. Tiempo, aspecto y la interpretación de los verbos modales en español. *Linguística ALFAL* 17: 9–44.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar*, Vol. 2. Stanford CA: Stanford University Press.
- Langacker, Ronald W. 2013. Modals: Striving for control. In *English Modality: Core, Periphery and Evidentiality*, Juana I. Marín-Arrese, Marta Carretero, Jorge Arús & Johan van der Auwera (eds), 3–55. Berlin: Mouton de Gruyter. doi:10.1515/9783110286328.3
- Langacker, Ronald W. 2017. Evidentiality in Cognitive Grammar. In *Evidentiality Revisited: Cognitive Grammar, Functional and Discourse-pragmatic Perspectives* [Pragmatics & Beyond New Series 271], Juana I. Marín Arrese, Gerda Hassler & Marta Carretero (eds), 13–55. Amsterdam: John Benjamins. doi:10.1075/pbns.271.02lan

- Marín Arrese, Juana I. 2013. Stancetaking and inter/subjectivity in the Iraq inquiry: Blair vs. Brown. *English Modality: Core, Periphery and Evidentiality*, Juana I. Marín-Arrese, Marta Carretero, Jorge Arús & Johan van der Auwera (eds), 411–445. Berlin: Mouton de Gruyter. doi:10.1515/9783110286328.411
- Marín-Arrese, Juana I. 2017a. Stancetaking and inter/subjectivity in journalistic discourse: the engagement system revisited. *Evaluation in Media Discourse: European Perspectives*, Ruth Breeze & Inés Olza (eds), 21–48. Bern: Peter Lang.
- Marín-Arrese, Juana I. 2017b. Multifunctionality of evidential expressions in discourse domains and genres: evidence from cross-linguistic case studies. *Evidentiality Revisited: Cognitive Grammar, Functional and Discourse-pragmatic Perspectives* [Pragmatics & Beyond New Series 271], Juana I. Marín Arrese, Gerda Hassler & Marta Carretero (eds), 195–223. Amsterdam: John Benjamins. doi:10.1075/pbns.271.09mar
- Oliveira, Teresa. 2015. Between evidentiality and epistemic modality: The case of the future and the conditional in European Portuguese. *Evidentiality and the Semantics-Pragmatics Interface*, Bert Cornillie & Juana I. Marín-Arrese (eds). Special issue of *Belgium Journal of Linguistics* 29: 101–121.
- Palmer, Frank R. 1990[1979]. *Modality and the English Modals*, 2nd edn. London: Longman.
- Palmer, Frank R. 2001[1986]. *Mood and Modality*, 2nd edn. Cambridge: CUP. doi:10.1017/CBO9781139167178
- Peirce, Charles S. 1955. *Philosophical writings of Peirce*, Justus Buchler (ed.). New York NY: Dover.
- Plungian, Vladimir. 2001. The place of evidentiality within the universal grammatical space. *Journal of Pragmatics* 33: 349–357. doi:10.1016/S0378-2166(00)00006-0
- Sanders, José & Spooren, Wilbert. 1996. Subjectivity and certainty in epistemic modality: A study of Dutch epistemic modifiers. *Cognitive Linguistics* 7(3): 241–264. doi:10.1515/cogl.1996.7.3.241
- de Saussure, Louis. 2013. Perspectival interpretations of tenses. In *Time. Language, Cognition and Reality*, Kasia M. Jaszczolt & Louis de Saussure (eds), 46–69. Oxford: OUP. doi:10.1093/acprof:oso/9780199589876.003.0004
- de Saussure, Louis & Morency, Patrick. 2012. A cognitive-pragmatic view of the French epistemic future. *French Language Studies* 22: 207–223. doi:10.1017/S0959269511000445
- Smirnova, Anastasia. 2013. Evidentiality in Bulgarian: Temporality, epistemic modality, and information source. *Journal of Semantics* 30: 479–532. doi:10.1093/jos/ffs017
- Squartini, Mario. 2001. The internal structure of evidentiality in Romance. *Studies in Language* 25(2): 297–334. doi:10.1075/sl.25.2.055qu
- Squartini, Mario. 2004. Disentangling evidentiality and epistemic modality in Romance. *Lingua* 114: 873–895. doi:10.1016/S0024-3841(03)00064-0
- Squartini, Mario. 2008. Lexical vs. grammatical evidentiality in French and Italian. *Linguistics* 46(5): 917–947. doi:10.1515/LING.2008.030
- Sthioul, Bertrand. 1998. Temps verbaux et points de vue. In *Le temps des événements*, Jacques Moeschler, Jacques Jayez, Monika Kozłowska, Jean-Marc Luscher, Bertrand Sthioul & Louis de Saussure (eds), 197–220. Paris: Kimé.
- Sweetser, Eve. 1982. Root and epistemic modals: Causality in two worlds. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 8: 484–507.
- Talmy, Len. 1988. Force dynamics in language and cognition. *Cognitive Science* 12: 40–100. doi:10.1207/s15516709cog1201_2
- Tasmowski, Liliane. 2001. Questions au conditionnel. *Le conditionnel en français*, Patrick Dendale & Liliane Tasmowski (eds), 331–343. *Recherches Linguistiques* 25. Metz: Université de Metz.

- Tournadre, Nicolas & LaPolla, Randy J. 2014. Towards a new approach to evidentiality. Issues and directions for research. *Linguistics of the Tibeto-Burman Area* 37(2): 240–263.
doi:10.1075/ltba.37.2.04tou
- van der Auwera, Johan & Vladimir Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2(1): 79–124. doi:10.1515/lity.1998.2.1.79
- Vatrican, Axelle. 2015. Evidentiality and epistemic modality in rumor: Journalistic conditional in Spanish. *Evidentiality and the Semantics-Pragmatics Interface*, Bert Cornillie & Juana I. Marín-Arrese (eds). Special issue of Belgium Journal of Linguistics 29: 83–99.
- White, Peter R. R. 2012. Exploring the axiological workings of 'reporter voice' news stories – Attribution and attitudinal positioning. *Discourse, Context & Media* 1: 57–67.
doi:10.1016/j.dcm.2012.10.004

Expressing sources of information, knowledge and belief in English and Spanish informative financial texts

Marta Carretero and Yolanda Berdasco-Gancedo

Universidad Complutense de Madrid / Universidad a Distancia de Madrid
(UDIMA)

This chapter presents a qualitative and quantitative analysis of expressions of sources of information, knowledge and belief and their combination with grammaticalized devices of future time from the perspective of the present and the past ('will', 'would' and their Spanish equivalents) in a corpus of English and Spanish informative financial texts belonging to two levels of specialization. The qualitative analysis describes the expressions concerned and their use in the corpus. The quantitative analysis uncovers the influence of language and level of specialization on the relative frequency of different expressions of source of information, knowledge and belief, a tendency for future time devices to occur within their scope, and a positive relation between the relative frequencies of both kinds of expressions.

Keywords: financial texts, reporting, epistemic modality, evidentiality, future time

1. Introduction and structure of the chapter

This chapter focuses on the explicit expressions of sources of information, knowledge and belief found in a corpus of English and Spanish financial texts with informative content. The interest of this study lies in the important role of these expressions in this kind of texts, supposedly aimed at transmitting information in an objective way and without much evaluation on the part of the journalist. However, a linguistic analysis in some depth reveals that the transmission of information in financial texts is not as objective as it seems; as Martínez-Egido (2015) points out, these texts display expressions that covertly transmit ideology, such as

many of the devices analyzed here (some reporting verbs, epistemic modal and evidential expressions) together with clausal mood and information structure.¹

The label ‘ideology’ points to the editorial line of the newspaper, an issue that lies outside the scope of this chapter. Our approach concentrates on a related issue, namely the expression of sources of information, knowledge and belief by which the journalist modulates his/her commitment to the information that s/he transmits. Informative financial discourse has two features, which have motivated their choice for this study and unavoidably lead to the appearance of the journalist’s voice in the message. As will be stated below, these features have led to the formulation of three research hypotheses.

The first feature is that the information reported by the journalist is rarely first-hand; that is, s/he reports the information given by experts who analyze the reported phenomena or even make decisions about them. Given this situation, the journalist, not surprisingly, tends to modulate the commitment to the truth of what s/he reports. If there is no explicit expression concerning this commitment, it is assumed to be full, as in (1) where the reported facts are (presented as) assumed to be true. In other cases, personal commitment is not full; for example, in (2) the journalist uses indirect reported speech citing a source, ‘A leading BVI lawyer’, but does not specify his/her own commitment to the truth of the proposition; lack of journalist’s commitment is also expressed in (3), this time with the evidential adverb ‘reportedly’ (Celle 2009).

- (1) Negotiations focused on the creation of a eurozone Single Resolution Mechanism (SRM) with the powers to close failing banks combined with a new financial supervisor for the eurozone under the auspices of the European Central Bank. (Level 1, Text 3)
- (2) Britain’s offshore satellites, in particular the British Virgin Islands, known for their secretive shell companies, do well out of this. *A leading BVI lawyer says* that Russian clients make up 15–20% of his business. (Level 2, Text 8)
- (3) However, deal activity dropped by 22 percent last year to \$124bn: *reportedly* the worst year the industry has suffered since 2003 (not including the catastrophic year that was 2009). (Level 2, Text 17)

We believe that the expressions for modulating commitment are highly dependent on the readers to which the articles are directed, and hence will be influenced by the overall cultural context (UK versus Spain) and also by the level of specialization. Our first hypothesis, derived from this belief, is that the overall frequency of these

1. Obviously, many references such as White (2012) and Mañoso-Pacheco (2017) attest a high number of covert ways for communicating evaluation in news discourse.

expressions, as well as the specific frequency of expressions of different types, will vary according to both language and level of specialization.

The second feature of informative financial texts is that the information is far from always being about past or present facts: within the area of finances, readers are especially concerned about future and potential events. This anxiety is understandable, since finances are key to the well-being of the world or part of it. Some readers may also have professional or personal interests, since they have to make financial decisions about money or other kinds of goods. A number of titles of the articles analyzed here strongly point to the future, such as ‘Fears over Chinese slow-down after trade data disappoints’ or ‘Diversify to succeed: Bank of the West promotes international investment’. In other cases, the news are mainly about present or past facts, but also mention future implications of these facts. In sum, statements about the future are common in these texts. Such statements cannot be facts, since they have not happened or are not happening yet; that is to say, they are non-factual speculations, thus belonging to irrealis.² Some expressions, such as grammaticalized devices like the English modal auxiliary ‘will’ or the Spanish future tense confer an almost factual status to future events. However, the journalist may well combine these devices with others that qualify this commitment, as in (4), where ‘will’ lies within the scope of reported speech and of the epistemic expression ‘certainty’:

- (4) “[...] and the only *certainty* is that base rate *will* climb at some point,” says Hollingworth. (Level 1, Text 1)

This irrealis status of future events has led us to formulate another two hypotheses: the second hypothesis is that, since statements about the future are speculations stemming from someone’s mind and not reports of actual facts, they will often lie within the scope of expressions of information, knowledge and belief, and more concretely within the scope of those expressions that attribute the information to a different conceptualizer from the journalist, because the journalist will avoid making his/her own predictions since they would not match the desirable aura of objectivity described above. The third hypothesis is that a correlation across texts will be found between the number of cases of future ‘will’ and the Spanish future and the total number of expressions of source of information, knowledge and belief, because texts that refer to future events more often will be more speculative; by contrast, the total number of these expressions will be lower in texts more centered on the narration of past and present events. We also predict that the last two hypotheses will be valid across the languages and levels of specialization; in principle, we see no reasons for differences in this respect.

2. According to Givón (1984: 285), irrealis assertions are those that do not present propositions as facts, but as hypothetical, possible, desirable or uncertain states or events.

The last two hypotheses are to be extended to the use of ‘would’ and its Spanish correlate, the conditional tense, when they express future from the perspective of the past, as in (5), since the reported speaker/writer refers to events posterior to the time when the message was produced.

- (5) The IEA calculated the Government *would* need to slash spending by more than a quarter or impose significant tax hikes because official calculations had failed to factor in future pension and healthcare liabilities. (Level 1, Text 11)

In order to test these three hypotheses, the expressions of source of information, knowledge and belief have been subjected to a qualitative and quantitative analysis in a corpus of 80 financial informative texts totalling approximately 50,000 words, evenly divided into language (English-Spanish) and level of specialization: 40 texts are extracted from general media aiming to reach a broad readership, and the other 40 texts are aimed at knowledgeable readers in the area of financial and economic information. The quantitative analysis covers a comparison of the frequencies of the different expressions of our concern depending on language and level of specialization; a study of the cases in which ‘will’ and ‘would’ and their Spanish equivalents lie within the scope of expressions of information, knowledge and belief; and the degree of correlation between the total frequency of these expressions in each text and the frequency of ‘will’ and ‘would’.

This chapter is structured as follows: the corpus is described in Section 2; a qualitative analysis of the different expressions of source of knowledge and belief found in the corpus is presented in Section 3 along with a specification of the contexts in which they do not express sources of knowledge and belief and are therefore excluded from analysis; the cases of future ‘will’ and conditional ‘would’ as well as their Spanish equivalents are described in Section 4, in order to check the second and third hypotheses; the methodology for data gathering and the quantitative analyzes are explained in Section 5; the results are presented in Section 6, while Section 7 sums up the main conclusions and offers suggestions for further research.

2. Description of the corpus

The analysis has been carried out on 80 texts, 40 from non-specialized quality journals and 40 from specialized journals, also evenly divided into British English and Peninsular Spanish texts, totalling about 25,000 words for each language. For the sake of comparability, all the texts share the following characteristics:

- informative texts dealing with financial and macroeconomic issues.

- extracted from media webs, fully accessible except *Financial Times* and *The Economist*, which required registering.
- length of at least 500 words.
- published between 2013 and 2015.

The texts were assigned a level of specialization according to what is proposed by others (e.g. Fernández del Moral & Esteve 1996; Hoffman 1998; Cabré 1999). Accordingly, the texts from non-specialized journals belong to Level 1, since they are extracted from general media meant to reach a broad readership. The texts from specialized journals belong to Level 2, since they are addressed at people interested in financial and economic information, with previous knowledge about the area of specialization. The Level 1 texts have been extracted from the British newspapers *The Guardian*, *The Independent* and *The Times*, and the Spanish newspapers *El País*, *ABC* and *El Mundo*. The Level 2 texts have been extracted from the British newspapers *Financial Times*, *The Economist* and *World Finance*, and the Spanish newspapers *Cinco Días*, *Expansión*, *Finanzas.com* and *Banca15.com*. The identification number, source and title of the texts is specified in the Appendix.

Although defining the extension of a corpus can be difficult, especially with such a great population as this, we consider that the sample chosen is representative and that the results obtained can be useful for future research. Since it totals approximately 50,000 words, this corpus follows the prescriptions given by Fox (1999: 264):

Whereas general language corpora are recommended to be as large as possible, investigations into technical and professional languages have demonstrated the representativeness of small-domain specific corpora, totaling 20,000–30,000 running words [...], a consequence of the restrictions of specialized languages: predictable topics, limited syntactic structures of unusual distribution and specialized vocabulary.

The sample plays a crucial role in this kind of research, but there are other important elements. Biber (1993: 244) thus focuses on two other factors that he considers of more importance than the sample size itself: correct definition of the target population and due consideration of its variability. This is why we selected English and Spanish texts and stratified our sample: the texts were divided into two levels according to their degree of specialization. These two levels are defined a priori by the media where they are published and the expected readers. The advantages of doing a stratified classification are also mentioned by Biber (1993: 244):

stratified samples are almost always more representative than non-stratified samples (and they are never less representative). This is because identified strata can

be fully represented (100% sampling) in the proportions desired, rather than depending on random selection techniques. In statistical terms, the between-group variance is typically larger than within-group variance, and thus a sample that forces representation across identifiable groups will be more representative overall.

This even selection of texts according to level of specialization has enabled us to carry out the quantitative research described in the sections that follow.

The texts belong to the corpus analyzed previously in Berdasco-Gancedo (2016), where the level of specialization was shown to be reflected in a number of linguistic differences, especially concerning syntax, lexical density and reported speech. The present research complements this previous work through the description of the influence of the level of specialization on the expression of source of information, knowledge and belief.

3. Expressions of source of information, knowledge and belief: A qualitative analysis

This section is an account of the different categories of expressions of source of information, knowledge and belief analyzed here, which are divided into two large subtypes: expressions of epistemicity, a category that includes epistemic modality and evidentiality, and expressions of reported speech. The subsections start with conceptual issues about each category, so as to clarify its scope, and continue with a qualitative analysis of each of the expressions in the texts, although occasionally quantitative data have been used in a non-systematic way when they were considered of interest to give precision to parts of the description.

3.1 Expressions of epistemicity

We adopt the term epistemicity from Boye (2012), who uses it as an umbrella category covering the categories of epistemic modality and evidentiality. Epistemicity may be defined as a qualification of the commitment to the information transmitted by a proposition; epistemic modality conveys this qualification by expressing the estimation of the chances for a proposition to be or become true (cf. Nuyts 2001: 21), as in (6), while evidentiality conveys it by expressing the kind, source and/or evaluation of evidence for or against the communicated proposition as in (7).

- (6) There are two things to bear in mind, though. First is that other ISA providers *may* come in with better offers before the 5 April deadline. (Level 2, Text 1)

- (7) Mr Flowers is on police bail having been filmed *allegedly* buying drugs
(Level 1, Text 2)

The scope of epistemic modal expressions and of evidential expressions is a proposition, that is, a piece of information that has truth value. Some expressions are evidential if their scope is a proposition, but not if it is a state of affairs. According to Boye (2012: 192), the difference between both lies in that “states of affairs can be said to occur, whereas propositions can be said to have a truth value”. Some expressions can have both states of affairs and propositions within their scope. For example, the verb ‘suggest’ is evidential in (8), since it qualifies the commitment to the proposition expressed by the span “there... had thought”, but not in (9), where it expresses a state of affairs, that is, the bringing about of the action indicated by the span “including... 24 hours”:

- (8) “Taken in isolation, the developments on the wage front *suggest* to me that there has been more spare capacity in the labor market than we previously had thought”, he said (Level 1, Text 11)
- (9) Mr Constâncio *suggested* including a fast-track process to enable decisions to be made within 24 hours. (Level 2, Text 2)

We follow Boye (2012: 183–184) in considering scope as a property of linguistic meanings rather than linguistic expressions. Accordingly, an expression can be epistemic modal or evidential even if its explicit scope is not a full clause, provided that its implicit scope is a proposition. For example, in (10) the noun ‘prospect’ is epistemic modal although its syntactic scope is the Prepositional Phrase ‘of higher yields in rich countries’, since there is an implicit proposition qualified, namely ‘the yields will be higher in rich countries’:

- (10) When in May the Federal Reserve began discussing plans to scale back its asset purchases, the *prospect* of higher yields in rich countries made investors reluctant to pour more money into emerging economies. (Level 2, Text 5)

Epistemic modality and evidentiality will be considered as functional-conceptual substance domains, a label coded in Boye & Harder (2009: 10), according to which all the expressions with epistemic modal and evidential meanings will be included independently of their morphology and syntax, “without interference from structural criteria associated with different forms of coding”. This label will be used here in a broad sense, encompassing devices without restriction to certain grammatical categories. The resulting scope for each category will be explained in Sections 3.1.1 and 3.1.2.

3.1.1 *Epistemic modal expressions*

Epistemic modality, which was defined above as the estimation of the chances for a proposition to be or become true, is realized by expressions of possibility, probability and certainty. Typical English expressions are the modal auxiliaries in their epistemic meaning, adjectives such as ‘possible’, ‘probable’ or ‘certain’ and the corresponding adverbs, and nouns such as ‘possibility’, ‘probability’ and ‘certainty’. We also include complement-taking predicates expressing knowledge such as ‘know’, and predicates expressing belief (which means lack of total certainty), such as ‘think’ or ‘believe’. All the expressions cited in this paragraph have narrow Spanish equivalents.

The scope of epistemic modality also comprises the so-called apprehensional-epistemic modality in Lichtenberk (1995: 293), which “has to do with the speaker’s degree of certainty about the factual status of a proposition and also with his or her attitude concerning the desirability of the situation included in the clause”.³ That is to say, these expressions share the semantic element of belief with verbs such as ‘think’ and ‘believe’, which motivates their consideration as epistemic, but also have a semantic element of desirability or undesirability of the truth of the proposition. Some examples are verbs such as ‘hope’, ‘expect’ or ‘fear’ and their Spanish equivalents *esperar* (‘hope’) and *temer* (‘fear’), adjectives such as ‘hopeful’, ‘promising’ or ‘dangerous’ and their respective Spanish equivalents *esperanzador*, *prometedor* or *peligroso*,⁴ and nouns such as ‘hope’, ‘fear’, ‘danger’ or ‘risk’ and their Spanish equivalents *esperanza*, *miedo/temor*, *peligro* and *riesgo*, when they are used as epistemic modal qualifications about a proposition.

Once the scope of epistemic modality has been determined, we specify the threefold distinction that will be made for the qualitative and quantitative analysis, depending on the journalist’s position with regard to the epistemic modal qualification.

The first category includes those cases in which the journalist him/herself is the formulator of the epistemic modal qualification and therefore claims responsibility for it, as in (6) above. The second category covers thought attribution, by which the journalist attributes a belief to someone else (11); this category also comprises epistemic modal expressions under the scope of verbs of thought attribution when the modal qualification is attributed to the same person: for example, this is the case of ‘more likely’ in (12), since this qualification is attributed by the journalist to the referents of ‘many experts’ in the same way as ‘think’.

3. See also Lavid et al. (2016: 12; 2017: 248).

4. Spanish adjectives inflect for gender and number. Throughout the chapter, the cited form is the masculine singular.

- (11) With China's stock of credit soaring to about 200 per cent of gross domestic product, from 130 per cent just five years ago, *analysts believe* the government is trying to encourage companies and investors to reduce their debt loads. (Level 2, Text 3)
- (12) *many experts think* taxes are *more likely* to rise than fall after the next election to protect public services and keep the deficit falling. (Level 2, Text 14)

The third category comprises the epistemic expressions that lie within the scope of direct or indirect reported speech, which goes one step further in the writer's distancing from the responsibility of the epistemic modal qualification: the journalist does not directly attribute the qualification to someone else, but claims that someone else has actually made this qualification. In other words, the qualification is attributed to another formulator, which may be a person (13), an institution (14) or even a report or another kind of written document.

- (13) "We have never been in a situation like this before. It is *quite possible* that we will not find our way through without serious social breakdown and/or mass emigration of the most mobile and productive people," said Mr Booth. (Level 1, Text 11)
- (14) The IMF said low interest rates *might* spur risk-taking behaviour by investors who demand a higher rate of return on their investments. (Level 1, Text 15)

We have classified the epistemic modal expressions in the first category when the journalist conceals his/her role of formulator of this qualification or, in Marín-Arrese's words (2011: 44), "there is mystification in his/her responsibility for the assessment". The qualification is formulated in an impersonal way, but the journalist is implicitly included in this impersonalization. Examples are 'the prospect' in (10) above, and 'it is thought' and 'could' in (15):

- (15) *It is thought* that the unprecedented crackdown *could* put all but the three biggest players out of business. (Level 1, Text 5)

This threefold classification will allow us to see differences of frequencies of expressions with different degrees of responsibility on the part of the journalist depending on level of specialization and language, thus providing a way to check the first hypothesis. It is also relevant to the third hypothesis, which predicts that 'will' and 'would' and their Spanish equivalents will be found to be within the scope of expressions by which the journalist transfers responsibility for the assertion of a future state or event to someone else; in other words, 'will' and 'would' will be often found within the scope of direct and indirect reported speech or of thought attribution

We now proceed to a qualitative analysis of the epistemic modal expressions found in the corpus. The two languages display a similar layout in terms of syntactic categories and meanings:

1. Some English modal auxiliaries, such as ‘may’, ‘might’ and ‘could’, and less frequently ‘should’. The Spanish closest equivalent, the periphrasis *poder* followed by an infinitive, was also common. However, no epistemic occurrences were found of English ‘must’ nor of its Spanish equivalent *deber / deber de*; both were always used with the deontic meaning of obligation.
2. Epistemic modal lexical verbs (also called epistemic propositional attitude verbs), such as ‘think’, ‘believe’, ‘hope’ or ‘fear’ and their respective Spanish equivalents *pensar*, *creer*, *esperar* and *temer*. These verbs are nearly always found within the scope of either reported speech or thought attribution, as is the case of the 9 occurrences of ‘believe’ in the English corpus. This fact agrees with an overall scarcity of subjective expressions such as ‘I think’ or ‘I believe’ to express qualifications having the journalist as conceptualizer, which is not surprising, since these expressions would clash with the aura of objectivity that these financial texts are supposed to transmit.
3. Adjectives, such as English ‘likely’, ‘possible’ and ‘plausible’. In Spanish, epistemic modal adjectives are uncommon, although there are cases of adjectives such as *posible* (‘possible’) or *cierto* (‘certain’).
4. Adverbs, such as English ‘maybe’, ‘perhaps’, ‘possibly’, ‘probably’ or ‘certainly’, and Spanish *probablemente* (‘probably’), *previsiblemente* (‘predictably’) and the adverbial *sin duda* (‘no doubt’). These adverbs occur only once or twice in the corpus, except for ‘probably’, which occurs 7 times – in 5 of them within reported speech. No occurrences were found of Spanish *quizá/quizás* (‘perhaps’), *tal vez* (‘perhaps’) or *seguramente* (‘certainly’), the reason probably being that they sound too informal for this kind of text.
5. Nouns. Many epistemic modal nouns occur in both languages, some examples being English ‘assumption’, ‘danger’, ‘risk’, ‘doubt’ (as in (16)), ‘fear’, ‘hope’, ‘expectation’, ‘prospect’ and ‘warning’, and Spanish *amenaza* (‘threat’), *incertidumbre* (‘uncertainty’), *riesgo* (‘risk’) or *previsión* (‘forecast’).

- (16) The EU agrees ‘banking union’ but *doubts* remain over the detail of who decides and pays for bank restructuring or closure. (Level 1, Text 3)

3.1.2 Evidential expressions

We define evidentiality as the kind, source and/or evaluation of evidence for or against the communicated proposition. It will be treated from a functional-conceptual perspective, against much of previous literature, which has mostly concentrated on grammatical markers of evidentiality and more lexicalized expressions such as adverbs or verbs of perception or cognition (Aikhenvald 2004; Celle 2009; Hennemann 2012; Ruskan 2015; Usonienė & Šinkūnienė 2013). Before describing our scope, reference must be made to Anderson’s (1986: 274–275) proposal for

the scope of evidentiality in terms of four criteria: the first states that “[e]videntials show the kind of justification for a factual claim which is available to the person making that claim [...]”; the second, that evidentials cannot be the main predication of the clause; the third, that evidentials must indicate evidence as their primary meaning, not only as a pragmatic inference; and the fourth, that evidentials should be “inflections, clitics, or other free syntactic elements (not compounds or derivational forms)”. We acknowledge that studies on evidentiality in this sense have shed light on its status as a grammatical and/or semantic category and its ways of expression in a wide range of languages, which has meant an advance for typological studies and for theoretical and descriptive linguistics in general. However, we have opted for a scope encompassing all the expressions that meet Anderson’s (1986) first criterion, since we believe that the joint consideration of all the expressions which share the function of justifying factual claims sheds more light on the journalists’ use of such justifications for transmitting information. That is to say, all the expressions that indicate kind or source of evidence are included. For example, we include expressions such as the italicized stretch in (17), which lies outside evidentiality according to Anderson’s second and fourth criteria; the inclusion is based on the grounds that it indicates the source of evidence in favor of the proposition expressed by the stretch ‘turnover... \$416bn’, thus resembling other expressions more standardly considered as evidentials such as ‘according to’ (Wiemer & Stathi 2010: 282) or even ‘reportedly’, which could be used as paraphrases.

- (17) *WTO statistics showed that* – while UAE foreign trade decreased by 20 percent from 2008 to 2009 – by last year, turnover had risen to over \$520bn, well above the pre-crisis figure of \$416bn. (Level 2, Text 1)

As for the expressions that evaluate the evidence in terms of speaker/writer’s commitment to the truth of the proposition, a number of authors consider that they should be classified as epistemic. This is the case of De Haan (2005: 380) and of Wiemer & Stathi (2010: 276), who consider (18), an example as expressing degree of certainty, thus being epistemic:

- (18) Mr President, *clearly* the majority of Members have decided to limit their working time tonight, but never mind!
[cited from Simon-Vandenberg and Aijmer (2007: 167)]

However, we view ‘clearly’ as expressing high commitment to the truth of the proposition by evaluating the evidence, in contrast to other adverbs of certainty such as ‘certainly’ or ‘undoubtedly’, which express a similar degree of commitment without indicating its being grounded by evidence. According to this characterization, ‘clear(ly)’ and similar expressions such as ‘obvious(ly)’ and ‘evident(ly)’ could be

described as having the epistemic feature of expressing the commitment to the truth of the proposition and the evidential feature of grounding it on evidence. Therefore, they could be considered as ‘epistential’ (Faller 2002; Lampert 2015). However, we have integrated these expressions with evidentials, in order to avoid the introduction of an extra category for a limited number of expressions.

Evidentiality is commonly divided into three basic types according to the domain of evidence on which the evidential qualification is based (see, e.g., Marín-Arrese 2013: 418–422): perceptual (evidence derived from sensory perceptions), cognitive (evidence derived from inferences triggered with the aid of personal knowledge of the world) and communicative (evidence derived from reading or hearing linguistic messages). Not surprisingly, the most frequent type of evidentiality found in the corpus is the communicative. Communicative evidentiality may be divided into reportative evidentiality, which is based on concrete communicative sources, and hearsay evidentiality, which is based on rumor or extended opinion. Even though some cases of hearsay evidentiality may be found in the corpus, such as (7) above, reportative evidentiality is much more frequent.

As stated in Chojnicka (2012: 173), the distinction between reportative evidentiality and reported speech (covered in Section 3.2) is not easy to draw:

Reportive [sic] evidentiality is concerned with marking information as coming from another speaker(s), i.e. indicating that the current speaker has not witnessed the event him/herself but has merely heard about it from other sources. [...]

Reported speech, on the other hand, brings together tools and devices used for attributing knowledge to another speaker. It is usually divided into direct and indirect speech, which differs in the degree to which the original speaker’s perspective – or voice – is preserved [...]

We follow Chojnicka’s criterion, thus classifying linguistic expressions depending on the journalist’s main aim: above all, reported speech expressions attribute the information to a concrete source, and reportative evidential expressions qualify the commitment to the information as not directly witnessed, but heard or read from other sources. According to this criterion, expressions that do not specify the source undoubtedly belong to reportative evidentiality. This is the case of adjectives such as English ‘apparent’ and adverbs such as English ‘allegedly’, ‘apparently’ and ‘reportedly’. In Spanish, no examples were found of common evidential adverbs such as *aparentemente* (‘apparently’) or *supuestamente* (‘supposedly’). We also consider as cases of evidentiality those in which the source of information is cited but the journalist’s attitude about what s/he communicates is clear, as in (19), where *tal como* (‘just as’) indicates that the journalist subscribes to the information obtained by the source ‘Gortázar’:

- (19) *Como decíamos y tal como señaló Gortázar, en el total del crédito no se observarán grandes variaciones.* (Level 2, Text 18)
 ‘As we said and just as Gortázar signalled, in the total of the credit great variations will not be observed.’

With regard to ‘according to’ and the Spanish equivalents *de acuerdo con*, *según* and *como* followed by the source of information, we include them within reportative evidentiality along the lines of Chojnicka (2012: 174–175) even if they do not specify the journalist’s qualification, on the grounds that the source of information is part of the same proposition as the content of the report, not in a separate proposition as in the expressions considered as reporting in 3.2. An example is (20):

- (20) Households will be offered a child-care credit worth up to £2,000, helping as many as 2m families, **according to the government.** (Level 2, Text 7)

Evidentiality may also be expressed by lexical verbs which do not specifically communicate a verbal message, such as English ‘mean’ (with the meaning of ‘imply’), ‘seem’, ‘appear’, ‘look’, ‘show’, ‘demonstrate’ (with the meaning of ‘show’) and ‘suggest’ (with the meaning of producing an idea in the mind), and Spanish *parecer* (‘seem, appear’), *demostrar/mostrar* (‘show’), *indicar* (‘indicate’), *significar* (‘mean’) or *suponer* (‘mean’). These verbs can also express perceptual or cognitive evidentiality, but in our corpus they most often express communicative evidentiality, as is the case of ‘showed’ in (17).

Evidential expressions are subjected to a threefold classification as well: the first type includes qualifications formulated by the journalist him/herself, as in (7); the second comprises those cases in which evidential qualifications are attributed to someone else, in a parallel way to thought attribution, as with ‘seem’ in (21); the third type covers evidentials embedded in direct and indirect reported speech, as is the case of ‘weaker than expected wage growth suggested’ in (22). These differences will also be reflected in the quantitative analysis.

- (21) Yet the great mass of earners have heard this before. **For them** the national political discourse *seems* to take place at the extremities of British life. (Level 1, Text 16)
- (22) However, **he said that weaker than expected wage growth suggested** there was still considerable a significant amount of slack in the economy to be used up before a rise would be necessary. (Level 1, Text 18)

Apart from the predominance of communicative evidentiality stated above, a feature worth pointing out about the evidential expressions found in the corpus is the high frequency of evidential nouns, such as ‘sign’ in (23), which is parallel to the

high frequency of epistemic modal nouns (see Section 3.1.1). Examples of nouns occurring in the corpus are English ‘evidence’, ‘indication’, ‘proof’, ‘sign’, ‘clue’ and ‘allegation’, and Spanish *prueba* (‘proof’), *indicación* (‘indication’), *indicio* (‘indication’) and *señal* (‘sign, signal’):⁵

- (23) The proportion of Britons claiming jobseekers’ allowance fell to 3.5% in February: a further **sign** of labor-market strength and a dose of fiscal good news, since it reduces pressure on benefits. (Level 2, Text 7)

An expression found only in the Spanish corpus is the rumor conditional (Vatrican 2015), which indicates acknowledgement of information from another source, which may be explicit or not, without expressing degree of commitment, as in (24):

- (24) *Según fuentes conocedoras de lo establecido a nivel global, para los dos primeros habría establecido un nivel mínimo de capital exigible equivalente al 9% de los activos ponderados por riesgo y para el resto habría una horquilla entre el 7% y el 8%, según la base de comparación.* (Level 2, Text 16)
 ‘According to sources knowledgeable of what is established at a global level, for the first two a minimal level of demandable capital [lit would have been] established equivalent to 9% of risk-weighted assets, and for the rest there [lit. would be] a fork between 7 and 8% depending on the base of comparison.’

3.2 Reporting expressions

Reporting expressions are those by which the journalist explicitly attributes the information to another source different from him/herself. This type includes direct and indirect reported speech. Some authors, such as Li (1986: 41) consider that reported speech (“direct quote and indirect quote”) is included within evidentiality; however, we will consider them as separate categories following others (e.g. Chojnicka 2012; Wiemer & Stathi 2010; Boye 2012).

The most frequent verb used in the corpus for introducing the reported information is by far ‘say’ and its Spanish equivalent *decir*, which do not give indications of the journalist’s commitment to the information. Other less common verbs of this kind are English ‘tell’ or ‘note’ and Spanish *apuntar* (‘note’) or *exponer* (‘set forth’). However, these verbs are often used with sources of information whose mention has a conversational implicature of alignment with the cited source by conferring legitimacy to the information transmitted (Smirnova 2009: 98), most often persons or institutions with a political and financial high status, as in (25). Sources are sometimes inanimate, such as reports or other documents (26); more rarely,

5. For a detailed study of evidential nouns, see Carretero (2016).

they are indefinite, in which case they are often legitimized: as in (27), where two anonymous people are characterized as ‘close to the situation’.

- (25) **Phil Orford, the chief executive of the Forum of Private Business, said:** “Without more lending to small businesses, their growth and the country’s growth will remain slow. Politicians of all parties need to have a hard look at the lending sector.” (Level 1, Text 7)
- (26) “Without significant changes to spending levels, huge sacrifices will have to be made by future generations either through significantly higher taxes or reduced benefits,” **the report said.** (Level 1, Text 11)
- (27) The DoJ is focusing on sales staff rather than traders in its criminal investigation as it scrutinizes allegations of fraud in trading commissions (“mark-ups”) charged to clients, **two people close to the situation have said.** (Level 2, Text 12)

In other cases, the journalist resorts to verbs that have an evaluative meaning which, as stated in Martínez Egido (2015: 11), are covert transmitters of ideology. Some verbs, such as English ‘assure’, ‘explain’, ‘highlight’ or ‘insist’ and Spanish *destacar* (‘highlight’) *recordar* (‘remind’), or *señalar* (signal), have a conversational implicature of journalist’s alignment with the quoted source of information, which triggers a conversational implicature of journalist’s higher degree of commitment to the truth of the proposition in comparison to the neutral verbs mentioned in the previous paragraph. The meaning or implicature of undesirable or compromising information is often associated with high journalist’s commitment (otherwise the information would probably not have been communicated). This is the case of English ‘warn’ and its Spanish equivalent *advertir*, with 11 and 7 occurrences in the corpora respectively. Other verbs of this kind are English ‘acknowledge’, ‘admit’, ‘disclose’ and ‘reveal’ and their Spanish equivalents *reconocer*, *admitir* and *desvelar*.

By contrast, other verbs such as English ‘argue’ in (28), ‘claim’ or ‘suggest’ and Spanish *argumentar* (‘argue’) indicate weaker (or even lack of) alignment with the reported source, thus downtoning the journalist’s commitment to its truth.⁶

- (28) So now is exactly the right moment for the fund to sell with no fear of triggering a broader sell-off, **argues Takatoshi Ito, the chairman of the government-backed panel on the GPIF.** (Level 2, Text 6)

6. Some of the verbs mentioned in this section, as well other expressions, are treated in White (2012) from a dialogistic perspective, which differs from the truth functional account presented here in that emphasis is not laid on commitment to the truth of the proposition, but rather on dialogical engagement with prior utterances of the same topic and potential responses (White 2012: 61).

Other verbs convey emotional meanings; the evaluation is sometimes directed to the attitude of the source of information, as with English ‘brag’ and its Spanish equivalent *presumir*, or else the source directs it towards the content transmitted, as with English ‘praise’ and ‘accuse’ and Spanish *denunciar* (‘denounce’) or *ensalzar* (‘praise’).

Apart from verbs, noun phrases or prepositional phrases occur occasionally as reporting expressions, as in (29)–(30):

- (29) The **pledge** of Antonis Samaras, the Greek prime minister, that 2014 would be the year of economic recovery when Greece left its international rescue programme and returned to growth was met with disbelief, if not disdain, by the country’s business community (Level 1, Text 13)
- (30) *hasta que se estableció una cifra razonable* (‘*muy conservadora*’ **en palabras del presidente de Bankia, José Ignacio Goirigolzarri**) (Level 2, Text 10)
 (‘until a reasonable figure was established (“very conservative” in words of Bankia’s president, José Ignacio Goirigolzarri)).

The reported information is mostly expressed by clauses, although it may be realized by other kinds of syntactic units such as noun phrases or prepositional phrases, as in (31), which contains one of each.

- (31) **The government announced** a broad de-regulation of the private pensions market that **Johnson described** as “one of the biggest shakeups in the tax treatment of private pensions we have seen in a long time”. (Level 1, Text 3)

3.3 Contexts of exclusion

Many of the expressions mentioned above are not expressions of sources of knowledge and belief in some contexts: for example, epistemic and evidential expressions are excluded if they are within the scope of irrealis contexts (see Note 3). In these cases, they do not express actual commitment to the truth of a proposition; instead, they express qualifications that are not yet made but might arise given certain circumstances. For example, the adjective ‘likely’ is excluded in (32), since it does not express an epistemic modal qualification actually made by the journalist or a cited source; instead, it refers to hypothetical estimations that brokers will have to make, as a rule, when asking for a loan. In (33), the modal ‘could’ is included in the analysis, since it expresses an epistemic qualification actually made by the cited source and applicable if the situation expressed in the if-clause becomes true; by contrast, ‘fears’ is excluded, since this qualification is not yet made but could occur given certain circumstances.

- (32) The rules, which come into force on January 2, 2015, require that brokers first make clear to customers that they are dealing with a credit broker, what its legal name is, the likely size of the fee and how that fee will be payable.
(Level 2, Text 19)
- (33) If the movement gained momentum, managing to split apart from Spain, this “would set a precedent and **could** quickly revive fears of at least a partial break-up of the euro”.
(Level 1, Text 20)

A number of epistemic and evidential expressions are polysemous and therefore excluded when they have other meanings. This is the case of the modal auxiliaries and the Spanish equivalent periphrases when they express deontic or dynamic modality, of ‘clear’ and ‘clearly’ with a meaning of manner, and of verbs such as ‘suggest’ when they propose a course of action (see 3.1), and of ‘think’ when it expresses an opinion, which is non-verifiable and hence cannot be subjected to an epistemic qualification (34). The verb ‘see’ was also excluded when it refers to physical vision. A case was found of the verb ‘believe’ where it expresses mirativity (i.e. disagreement with previous expectations) instead of epistemic modality and was therefore excluded (35):

- (34) I don’t think it’s right that more than half a million British expatriate pensioners – including many war heroes – are being treated so shabbily by the Government.
(Level 1, Text 19)
- (35) “I can’t believe that this is the country where I grew up. For my generation it was never meant to be this way”.
(Level 1, Text 11)

Verbs of saying are not classified within the ‘reporting’ categories when they report directive speech acts, as in (36). In these cases, the information transmitted is a proposal of a course of action, the unit within the scope of the verb is a state of affairs and not a proposition (see 3.1.2), and, in Searle’s (1976) words, the direction of fit is world-to-word, not word-to-world:

- (36) The struggling Monte dei Paschi di Siena has the biggest shortfall and must raise €2.1bn. Portugal’s Banco Comercial has been told to increase its capital by €1.15bn. There are two lenders from Belgium and two from Slovenia on the list.
(Level 1, Text 15)

4. Expressions of future time: English ‘will’ and ‘would’ and their Spanish equivalents

The English modal auxiliary ‘will’ and the Spanish synthetic future, examples of which are (37)–(38), respectively, are the most grammaticalized means for expressing future time in the respective languages:

- (37) Consumers **will** have a 14-day right of cancellation where credit broking contracts are entered into online. (Level 2, Text 19)
- (38) *Las entidades locales recibirán 3.700 millones de euros menos en 2015 y la protección social sufrirá un recorte de 6.400 millones.* (Level 1, Text 2)
 ‘Local entities will receive 3,700 million euros less in 2015, and social protection will undergo a cut of 6,400 million.’

With regard to ‘will’ expressing future time, there is no agreement in the literature about whether it has modal meaning or should instead be considered as a factual pure future expression (see Lyons 1977; Larreya 1984; Coates 1983; Huddleston 1995, among many others). We believe that both positions are partially right: on the one hand, these grammaticalized devices could be considered as non-modal on the grounds that they conceive future states or events as if they were facts. On the other, unlike past and present events, future events cannot possibly be actual facts, so that they belong to the irrealis domain (see Section 1); ‘will’ may be seen as a device that explicitly indicates this status, in contrast to the Simple Present, as in ‘The train leaves at 9.30 tonight’, which treats future events on a par with present facts. In this sense, future ‘will’ may well be considered as a device for expressing epistemic modality.⁷ As for the Spanish future, even if its role is similar to that of English ‘will’, it is treated in Spanish grammars as a tense rather than a modal expression, the reasons being mainly formal (it is a bound inflectional suffix).⁸

In any case, it is precisely this non-factual status of future events that has triggered the second hypothesis stated in Section 1, namely that ‘will’ and the Spanish future tense are likely to be within the scope of the expressions of source of information, knowledge and belief described above, especially those by which the journalist attributes the information to someone else, and also the third hypothesis, namely

7. ‘Will’ can also express the modal meaning of volition, into which we cannot delve for reasons of space.

8. ‘Will’ and the Spanish future tense can also express probability in present or past time, as in ‘Mary will be having lunch now’/*Mary estará almorzando ahora* or ‘Mary will have arrived by now’/*Mary habrá llegado ya*, but this use has not been found in the corpus in either language.

that its total number has a positive correlation with the total number of expressions of information, knowledge and belief. As a more objective way of testing the third hypothesis, ‘will’ and its Spanish equivalent the future tense have been considered as a category independent of epistemic modality.

As we stated in the Introduction, the correlation also includes ‘would’ and its Spanish equivalent, the conditional tense, when they express future time within the scope of past time, as in (5) above; in certain cases, the future time is embedded in a hypothetical context or, in other words, the occurrence of the future event signalled with ‘would’ is conditioned to the occurrence of another event, as in (39), where it appears within the scope of a conditional if-clause that expresses an event whose occurrence conditions that of the future event. However, future possibilities are not at issue in counterfactual occurrences of ‘would’, such as (40), which are therefore discarded; the Spanish rumor conditional mentioned in Section 3.1.2 is also excluded from this count, since it is an evidential expression.

- (39) “If you let the private sector create money, there is no control over the supply of this money;” says Mark Weisbrot, Co-Director of the Centre for Economic and Policy Research. “This would be, in some ways, a step backward to the days before there was central banking.[...]
- (Level 2, Text 15)
- (40) A former Soviet bloc state such as Bulgaria offering stability and solace to entrepreneurs such as Agorastos **would** have been inconceivable not that long ago when Greece, the Balkans’ only EU member, was attempting to cast itself as the regional economic leader.
- (Level 1, Text 6)

Admittedly, the restriction of the correlation study to ‘will’ and the Spanish future tense, due to practical limitations, has the shortcoming that it does not exhaust the ways of expressing future in English and Spanish. Examples of other devices are the periphrastic expressions ‘be going to’ in English and its Spanish equivalent *ir a* (both of which occur very rarely in the corpus); moreover, in certain cases the Spanish equivalent to ‘will’ and ‘would’ is not the future or the conditional tenses but the subjunctive mood, as in (41).

- (41) *Los expertos consultados por Europa Press esperan que el Santander mantenga su valor en Bolsa por encima de los 100.000 millones en el largo plazo*
- (Level 1, Text 19)
- ‘The experts consulted by Europa Press hope that Santander [bank] will maintain (literally ‘maintain-pres.subj.) its value in stock market’

5. The data gathering

The quantitative analysis of the expressions of sources of information, knowledge and belief in the selected texts was carried out by means of an automatic search that served to signal the expressions cited in Sections 3.1 and 3.2, as well as ‘will’ and ‘would’. This search was followed by a manual analysis which allowed us to discard the excluded cases and to include unexpected expressions, considering that the descriptive sections could not possibly exhaust the myriad of linguistic devices that can express epistemicity and reported speech. In fact, this procedure permitted us to find a few cases in which these meanings are expressed with creative metaphors. Two such cases are *fantasma* (‘phantom’) in (42) and *música* (‘music’) in (43):

- (42) *El Gobierno francés hizo ayer un ejercicio de equilibrismo en la presentación de sus cuentas de 2015 al presumir de responsabilidad presupuestaria y tratar a la vez de alejar el fantasma del rigor.* (Level 1, Text 2)

‘The French government did yesterday a balance exercise in the presentation of its 2015 accounts, since it boasted of budgetary responsibility and attempted at the same time to ward off the phantom of rigor’

- (43) *La prima de riesgo española se sitúa por debajo de los 90 puntos básicos y parece estabilizarse por debajo de la cota de los 100 puntos, encantada con la música que le llega desde Fráncfurt sobre la letra pequeña del “manguerazo” del BCE.* (Level 1, Text 11)

‘The Spanish risk Premium is situated under 90 basis points and seems to be stabilising below the quota of 100 points, enchanted with the music that it gets from Frankfurt about the small print of the “hosing down” by the European Central Bank.’

In (42), ‘the phantom of rigor’ could be paraphrased as ‘the possibility that there will be rigour’ or ‘the possibility that rigorous measures will be adopted’; these paraphrases make its epistemic modal value clear. Similarly, *música* in (43) is included as an evidential, since it qualifies an implicit proposition that could be paraphrased as ‘the information given in the small print of the “hosing down” by the European Central Bank is favorable to the interests of the Spanish economy’.

The occurrences of the expressions of information, knowledge and belief under analysis, once selected in the ways described above, were registered in a database according to the categories specified in Sections 3.1 and 3.2. The categories distinguished in the database are specified in Table 1.

Another database was made with the examples of ‘will’ and ‘would’ specified in Section 4; if they occurred within the scope of an expression of the categories distinguished in Table 3, the category of that expression was also registered. In order to detect the actual source of the information, the scope was considered from

Table 1. Categories of expressions registered in the database

-
1. Reported speech
 2. Epistemic modality with the journalist as conceptualizer
 3. Evidentiality with the journalist as conceptualizer
 4. Epistemic modality as thought attribution
 5. Evidentiality as thought attribution
 6. Epistemic modality within the scope of reported speech
 7. Evidentiality within the scope of reported speech
-

a semantic rather than a syntactic point of view. For example, in (44), the semantic scope of the evidential *según los cálculos gubernamentales* ‘according to Government calculations’ covers the following clause, thus affecting not only the verb *se producirá* ‘will take place’, but also *alcanzará* (‘will reach’) and *recortará* (‘will cut’):

- (44) *Este escenario se producirá, según los cálculos gubernamentales, en 2019. Este año alcanzará el 4,4% y se recortará una décima el que viene*

(Level 1, Text 2)

‘This scenario will take place, according to Government calculations, in 2019.

This year it [the deficit] will reach 4.4% and it will decrease one tenth next year’

The data gathered in the two databases were submitted to a quantitative analysis. For quantification, the texts were divided into four subcorpora following the criteria of language and level of specialization. We now turn to the results.

6. Results and discussion

In this section we present the results of the quantitative analysis, focusing on the extent to which they have confirmed or refuted the three hypotheses stated in Section 1.

6.1 Overall frequency of the expressions of information, knowledge and belief in the four subcorpora

The presentation of the results will start with the total and average frequency of the expressions of source of information, knowledge and belief of all the types in the four subcorpora resulting from separation according to language and level of specialization (see Table 2). The result show that the expressions are clearly most frequent in the English Level 1 sub-corpus, while the frequency for the other three groups is similar.

Table 2. Total number of expressions of source of information, knowledge and belief and average number of expressions per text

| | English Level 1 | Spanish Level 1 | English Level 2 | Spanish Level 2 |
|------------------|-----------------|-----------------|-----------------|-----------------|
| Total number | 481 | 253 | 342 | 228 |
| Average per text | 18.4 | 10.85 | 13.35 | 11.4 |

The overall frequency of the different types of expressions is specified in Table 3. Reported speech is by far the most frequent category, followed by evidentiality with the journalist as conceptualizer, epistemic modality as thought attribution and epistemic modality with the journalist as conceptualizer. Less frequent are the expressions of epistemic modality within the scope of reported speech. The least common categories are evidentiality within the scope of reported speech and evidentiality as thought attribution.

Table 3. Total number of expressions of source of information, knowledge and belief, divided into subtypes

| | |
|-------------------------------------------------------------|-----|
| 1. Reported speech | 367 |
| 2. Epistemic modality with the journalist as conceptualizer | 167 |
| 3. Evidentiality with the journalist as conceptualizer | 193 |
| 4. Epistemic modality as thought attribution | 169 |
| 5. Evidentiality as thought attribution | 18 |
| 6. Epistemic modality within the scope of reported speech | 114 |
| 7. Evidentiality within the scope of reported speech | 52 |

Table 4 specifies the distribution of the expressions of different kinds in the four subcorpora. The chi-square test shows that the variation in the distribution of the expressions through the subcorpora is significant, thus confirming the first hypothesis, namely that the distribution of the expressions varies with language and degree of specialization.

More specifically, we can see that the higher frequency of the expressions in the English texts of Level 1 is largely due to the category of reported speech, whose quantity in this subcorpus almost triples those of the other three subcorpora. That is to say, English Level 1 texts very often present the information as stemming from specific sources. This high frequency of reported speech is not surprisingly matched by a high frequency of epistemic modality and evidentiality within reported speech. A sample occurrence of an epistemic modal expression within reported speech is (45).

- (45) It is government gilt yields that determine annuity rates, and these are **unlikely** to rise until there is an uplift in the BoE base rate, **says Tim Gosden, head of annuities at Legal & General.** (Level 1, Text 1)

Table 4. Frequency of expressions of sources of knowledge and belief of all the types
 $\chi^2 = 123.9206$; $p = 7.58553 \times 10^{-18}$

| | English | | Spanish | | English | | Spanish | |
|-------------------------------------------------------------|---------|------|---------|------|---------|------|---------|------|
| | Level 1 | | Level 1 | | Level 2 | | Level 2 | |
| | Tot. | % | Tot. | % | Tot. | % | Tot. | % |
| 1. Reported speech | 170 | 46.2 | 64 | 29.5 | 66 | 24.7 | 67 | 29.4 |
| 2. Epistemic modality with the journalist as conceptualizer | 44 | 11.9 | 20 | 9.2 | 73 | 27.3 | 30 | 13.1 |
| 3. Evidentiality with the journalist as conceptualizer | 34 | 9.2 | 49 | 22.6 | 62 | 23.2 | 48 | 21 |
| 4. Epistemic modality as thought attribution | 38 | 10.3 | 44 | 20.3 | 40 | 14.9 | 47 | 20.6 |
| 5. Evidentiality as thought attribution | 2 | 0.5 | 4 | 1.8 | 5 | 1.9 | 7 | 3.0 |
| 6. Epistemic modality within the scope of reported speech | 57 | 15.4 | 23 | 10.6 | 16 | 6.0 | 18 | 7.9 |
| 7. Evidentiality within the scope of reported speech | 23 | 6.2 | 13 | 6.0 | 5 | 1.9 | 11 | 4.8 |
| TOTAL | 368 | 100 | 217 | 100 | 267 | 100 | 228 | 100 |

With regard to the other expressions, the distribution is quite even across the sub-corpora, with the exception that the English Level 2 has the highest number of expressions of both epistemic modality and evidentiality with the journalist as conceptualizer; this higher frequency hints that in English specialized financial publications the journalist has comparably more room to include his/her own speculations along with reported facts. (46) is an example of a Level 2 text that includes speculations of this kind, one containing an evidential expression ('seems') and another containing an epistemic modal expression ('probably'):

- (46) When George Osborne, the chancellor of the exchequer, rises to deliver his annual autumn statement on December 3rd, he could be forgiven a mild sense of satisfaction. British consumers are spending strongly. Firms, buoyed by consumers' confidence, are at last investing. Britain **seems** to be free from Europe's economic woes. Yet when it comes to his flagship economic policy – deficit reduction – Mr Osborne will **probably** deliver bad news: once again, he is missing his borrowing targets. (Level 2, Text 20)

The last two findings together permit to state that the variation of expressions across levels of specialization is clearly seen in the English texts: in Level 1, journalists have a strong tendency to attribute information to other sources, above all via reported speech, while in Level 2 they more readily make their own epistemic or evidential

qualifications, with the result that their voice sounds more 'expert' instead of the more 'orthodox' reporting found in Level 1. However, no comparable difference can be tracked from the results of the Spanish subcorpora, although Level 2 also displays a higher frequency of epistemic modal expressions with the journalist as conceptualizer (30 occurrences, versus 20 in Level 1).

6.2 Frequency of 'will/would' and their Spanish equivalents within the scope of other expressions of information, knowledge and belief

Table 5 specifies the distribution of the cases of 'will' and 'would' and their Spanish equivalents depending on their occurrence within the scope of an expression of source of information, knowledge and belief. The two English auxiliaries and their Spanish equivalents occur within the scope of such expressions in more than 50% per cent of the cases in all the subcorpora. The higher frequency of such cases for 'will' in Level 1 texts, in comparison to 'will' or the future in the other three subcorpora, is another consequence of the importance of reported speech in this subcorpus.

Table 5. 'Will', 'would', Spanish future and Spanish conditional: cases occurring and not occurring within the scope of an expression of source of information, knowledge and belief

| | | 'Will' | | Spanish future | | 'Would' | | Spanish conditional | |
|---------|------------------|--------|-------|----------------|-------|---------|-------|---------------------|-------|
| | | No. | % | No. | % | No. | % | No. | % |
| Level 1 | Within scope | 45 | 70.31 | 17 | 53.12 | 49 | 81.67 | 5 | 100.0 |
| | Not within scope | 19 | 29.69 | 15 | 46.88 | 11 | 18.33 | 0 | 0.00 |
| | Total | 64 | 100 | 32 | 100 | 60 | 100 | 5 | 100 |
| Level 2 | Within scope | 31 | 58.49 | 31 | 55.36 | 17 | 70.83 | 9 | 81.82 |
| | Not within scope | 22 | 41.51 | 25 | 44.64 | 7 | 29.17 | 2 | 18.18 |
| | Total | 53 | 100 | 56 | 100 | 24 | 100 | 11 | 100 |

Table 6 specifies the kind of expressions with the broader scope; for all the subcorpora, the majority of the expressions belong to categories where the journalist is not the conceptualizer, i.e. reported speech, epistemic and evidential expressions within reported speech, and thought attribution. These results may be considered as a confirmation of the second hypothesis. However, the percentages vary depending on the actual expression and the subcorpus. Not surprisingly, 'would' and the Spanish conditional are much more prone to occurring within the scope of the expressions of our concern than 'will' and the Spanish future, since the past perspective from

Table 6. ‘Will’, ‘would’, Spanish future and Spanish conditional: types of expressions under their scope

| | | ‘Will’ | | Spanish future | | ‘Would’ | | Spanish conditional | |
|---------|------------------------------------------------|--------|-------|----------------|-------|---------|-------|---------------------|-------|
| | | No. | % | No. | % | No. | % | No. | % |
| Level 1 | Reported speech | 30 | 66.67 | 8 | 47.06 | 37 | 75.51 | 3 | 60.00 |
| | Thought attribution and within reported speech | 7 | 15.56 | 2 | 11.76 | 9 | 18.37 | 2 | 40.00 |
| | Journalist as conceptualizer | 8 | 17.78 | 7 | 41.18 | 3 | 6.12 | 0 | 0.00 |
| | Total | 45 | 100 | 17 | 100 | 49 | 100 | 5 | 100 |
| Level 2 | Reported speech | 17 | 54.84 | 8 | 25.81 | 10 | 58.82 | 0 | 0.0 |
| | Thought attribution and within reported speech | 6 | 19.35 | 13 | 41.94 | 3 | 17.65 | 5 | 55.56 |
| | Journalist as conceptualizer | 8 | 25.81 | 10 | 32.26 | 4 | 23.53 | 4 | 44.44 |
| | Total | 31 | 100 | 31 | 100 | 17 | 100 | 9 | 100 |

which they express future is most often specified. In terms of absolute frequency, *would* is much more common in English Level 1 than in Level 2 texts, due to the higher frequency of reported speech in the first level. The Spanish conditional is scarce in both levels; in order to express a future perspective from a past report, other expressions are often preferred such as the future or even the simple present, as in (47). This is, in all probability, the reason why the Level 2 Spanish subcorpus has a higher percentage than the other types of occurrences of the future within the scope of an epistemic modal or evidential expression within reported speech or of an expression of thought attribution.

- (47) Y añadió que las perspectivas para 2015 son buenas (Level 1, Text 13)
 ‘And [he] added that the prospects for 2015 are good.’

6.3 Correlation between total number of expressions and the frequency of ‘will’/‘would’ and their Spanish equivalents

In this subsection we present an analysis of the relation between the frequency of ‘will’/‘would’ and that of all the expressions of information, knowledge and belief in the texts under analysis. Figures 1 to 4 represent both frequencies for each subcorpus. Concerning the English texts of Level 1 (Figure 1), we observe that the

two lines are far from parallel with the exception of the segments of texts 16 to 20. To this parallelism of the final segments we may add that texts 1, 4, 11, 14, 18 and 20 have a large number of expressions of source of information, knowledge and belief, and also more than 5 cases of ‘will’/‘would’. These texts tend to be speculative; for instance, Text 11 is titled ‘Bank of England behaving like an unreliable boyfriend, Say MPs’, and Text 20 is titled ‘Six triggers that could renew a eurozone crisis in 2015’. Conversely, the texts with fewer expressions of source of information, knowledge and belief, such as Texts 4, 5 and 6, and to a lesser extent Texts 10 and 17, also have few or no cases of ‘will’/‘would’. On the other hand, Texts 3, 12 and 15 contain as many occurrences of ‘will’/‘would’ as expressions of the other types (mostly expressions of reported speech), and consequently make the lines intersect. In its turn, text 7 has very few expressions of both kinds, which makes the lines converge. As suggested by its title, ‘Banks are ‘slowing growth’ by failing to lend Treasury cash to small businesses’, it consists almost exclusively of past and present events presented as true.

As for the Spanish texts of Level 1, Figure 2 shows that the correlation between the two variables is not strong either, even though, like in the English texts of Level 1, a relation may be found, since those texts having more expressions of source of information, knowledge and belief have two or more occurrences of the future or the conditional tenses (Texts 1, 2, 7, 8, 11 and 14), and conversely most of the texts that have no occurrences of the future or the conditional tenses also have few expressions of the other types (Texts 4, 5, 6, 17 and 20). A curious case is text 12, which contains a total of 7 expressions of which 6 are cases of the Spanish future: the text concerns requisites that banks have to meet in order to pay bonuses to their directives, thus being mostly concerned with obligations to be fulfilled. Qualification of knowledge and belief is then irrelevant. The future has a deontic implicature of obligation, as in *estará sometido a* (‘will be submitted to’) and is often combined with the modal auxiliary *deber* (‘have to’) as in *deberán limitar* (‘will have to limit’).

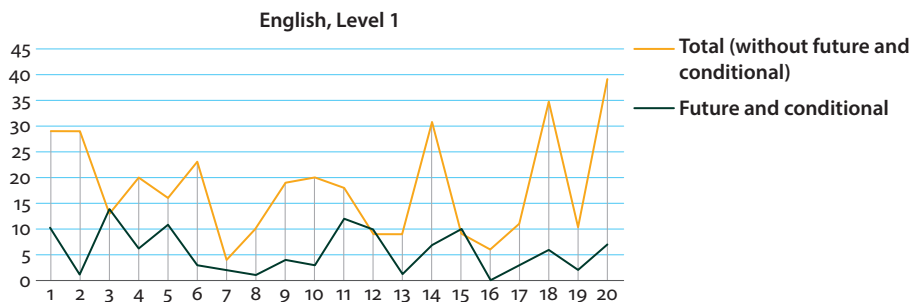


Figure 1. Frequency of ‘will’/‘would’ and of all expressions of sources of information, knowledge and belief in the English texts of level 1

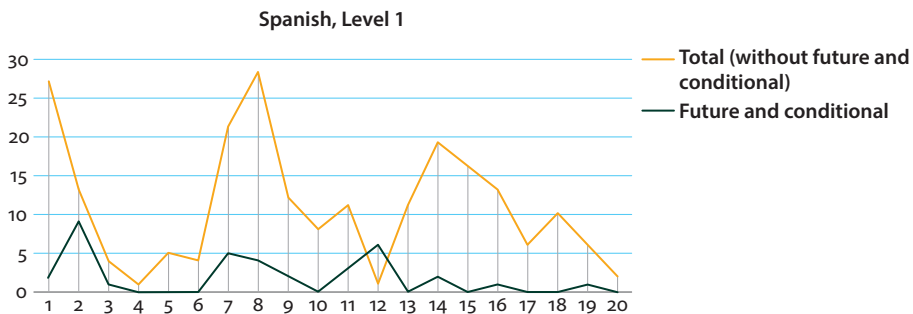


Figure 2. Frequency of ‘will’/‘would’ and of all expressions of sources of information, knowledge and belief in the Spanish texts of level 1

With regard to texts of Level 2 (Figures 3 and 4), the lines are closer to being parallel than for texts of Level 1 for both languages, especially the Spanish texts, where there is a clear parallelism between the segments corresponding to Texts 12 to 20. Those texts with fewer expressions of the two types tend to deal above all with past events, such as Text 17, titled with historic present “*España cobra por primera vez por emitir deuda a corto plazo*” (‘Spain charges for the first time for issuing short-term debt’), while texts with the two lines in higher position tend to be more focused on the future, as is the case of Text 20, titled “*La UE alerta: a más deuda, más ajuste fiscal en España*” (‘EU warns: to more debt, more fiscal adjustment in Spain’). We must also note that the future and conditional tenses are more common with these texts in comparison to Spanish Level 1, which means that reference to such non-factual events is more common. As for the English texts, a likely reason for the greater similarity of the two lines in Level 2 lies in the lower frequency of reported speech in comparison to Level 1, where one or more cases of ‘will’ and ‘would’ are often introduced by just one expression of reported speech. A case of ‘will’ typical of Level 2 may be seen in (46) above, where it occurs within the scope of ‘probably’ and the evidential expression ‘seems’ occurs in a close preceding clause.

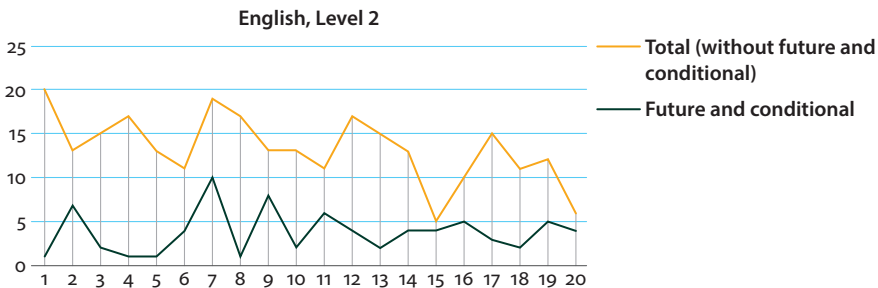


Figure 3. Frequency of ‘will’/‘would’ and of all expressions of sources of information, knowledge and belief in the English texts of level 2

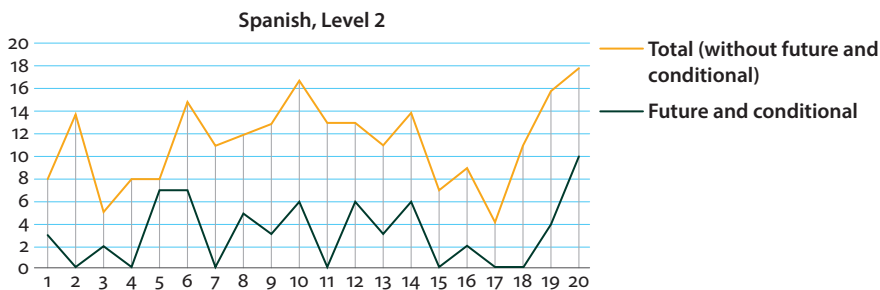


Figure 4. Frequency of ‘will’/‘would’ and of all expressions of sources of information, knowledge and belief in the Spanish texts of level 2

At a global level, we may then state then that the third hypothesis has been partially confirmed: in the texts analyzed, there is a relationship between the total number of expressions of information, knowledge and belief, and the total number of grammatical future devices, especially in the Level 2 texts in the two languages; however, this relation is not so narrow as to provoke straightforward parallelism between the totals of the two types of expressions.

7. Conclusions and suggestions for further research

This chapter has set forth a qualitative and quantitative analysis of the expressions of source of information, knowledge and belief, which qualify commitment to the information transmitted, on a corpus of 80 English and Spanish informative financial texts belonging to two levels of specialization. We characterized these texts in terms of two features that made us formulate three hypotheses: the first feature is that the information is rarely first hand, that is, nearly always provided by other personal sources; consequently, the journalist needs to qualify the commitment to the information communicated. For this reason, we formulated the first hypothesis, namely that the expressions under study would be frequent and also highly dependent on social and cultural factors, and consequently differences would be found in number and kind depending on language and level of specialization. This hypothesis has been confirmed, since the total number of the expressions under analysis in these texts may be considered as high. A number of common features were found across the four subcorpora:

- High frequency of reporting expressions, especially with the verb ‘say’ and its Spanish equivalent *decir*;

- Mystification of the journalists as conceptualizers of epistemic modal qualifications. This role is hidden by the use of intersubjective expressions; the few subjective expressions nearly always occur within the scope of reported speech, so that the qualification is attributed to the cited source and not to the journalist.
- High frequency of communicative evidential expressions.

In spite of these common tendencies throughout the texts, the chi-square test revealed that the variation in the distribution of subtypes across subcorpora is significant. The difference is more clearly seen in the two English subcorpora; English texts of Level 1 often present the information as stemming from other sources by means of reported speech, while English texts of Level 2 contain more speculations on the part of the journalist, which is reflected in a high frequency of epistemic modal and evidential expressions where *s/he* is the source of the qualification.

Another feature of these texts is frequent reference to future and potential events, since readers of financial news are interested in the future. The non-factuality of future events led us to formulate another two hypotheses. The second hypothesis was that ‘will’ and its Spanish equivalent the future tense, and also ‘would’ and the Spanish conditional, often lie within the scope of expressions of source of information, knowledge and belief, especially those subtypes in which the qualifier of the information is different from the journalist. The quantitative analysis uncovered that, in the four subcorpora, more than half the cases of ‘will’ and ‘would’ lay within the semantic scope of these expressions, and most of these expressions belonged to the predicted categories. Therefore, this hypothesis was also confirmed. The third hypothesis was that a cross-text correlation would be found between ‘will’/‘would’ and their Spanish equivalents and the total number of expressions of source of information, knowledge and belief. This hypothesis was partially disconfirmed, with the possible exception of Spanish Level 2 texts, since no strong relationship was found between the frequency of expressions of the two types. However, many cases were found of individual texts with either high or low quantity of expressions of the two types.

Further research might include similar analyses of the role of expressions of information, knowledge and belief and the correlation between these expressions and the devices for expressing future time in other text types such as news where the role of the future is theoretically less important, such as event news. The division into categories proposed here could also be refined: for example, reporting expressions could be subdivided into those that simply express lack of commitment to the truth of the information and those that emphasize or downtone this commitment (as in Mañoso-Pacheco 2017), epistemic modal expressions could be subdivided depending on the degree of probability expressed (as in, for instance,

Lavid et al. 2016), and evidential expressions could be divided into the perceptual, inferential and communicative subtypes described here. Or else the analysis could be deepened from the qualitative viewpoint, as in studies of the frequency and types of the expressions of source of knowledge and belief in texts of different discourse topics.

Acknowledgements

This research has been carried out as part of the EVIDISPRAG Project (reference number FFI2015-65474-P MINECO/FEDER), whose aim is to analyze evidential expressions in English and a number of European languages from a discourse-pragmatic perspective. We gratefully acknowledge the support provided by the Spanish Ministry of Science and Innovation and the European Regional Development Fund.

Appendix

This appendix specifies the identification number, source and title of the texts of the corpus. The translation of the titles of the Spanish texts are our own.

English texts, Level 1

| Text ID | Source | Title |
|---------|------------------------|---------------------------------------------------------------------------------------------------|
| 1 | <i>The Guardian</i> | Savers sense relief from rock-bottom interest rates. |
| 2 | <i>The Independent</i> | City advisers got £8.3m for Co-op deal but didn't check fatal loans. |
| 3 | <i>The Telegraph</i> | EU agrees 'general' principles of banking union. |
| 4 | <i>The Telegraph</i> | UK faces 'crippling' tax rises and spending cuts (to fund pensions and healthcare). |
| 5 | <i>The Independent</i> | How to get the best return from your cash. |
| 6 | <i>The Guardian</i> | Bulgaria benefits from weakness of Greek economy. |
| 7 | <i>The Guardian</i> | IFS says 2014 budget promises based on uncertain savings forecasts. |
| 8 | <i>The Guardian</i> | IMF warns slow economic recovery will keep interest rates at historic lows. |
| 9 | <i>The Telegraph</i> | Middle-earners want tax cuts, not more promises. |
| 10 | <i>The Independent</i> | Stop moaning about our bonuses: Barclays boss slaps down investors angry at bank's pay practices. |
| 11 | <i>The Guardian</i> | Bank of England behaving like an unreliable boyfriend, say MPs. |
| 12 | <i>The Independent</i> | Elderly people living in certain countries abroad are being denied pension rises. |
| 13 | <i>The Independent</i> | Italian banks' capital hole exposed by European nancial stress test. |

| Text ID | Source | Title |
|---------|----------------------|----------------------------------------------------------------------------------|
| 14 | <i>The Telegraph</i> | FCA price cap on payday lenders to 'put an end to spiralling debt'. |
| 15 | <i>The Times</i> | Juncker's critics hit out over €315bn plan to revive Europe. |
| 16 | <i>The Times</i> | Banks are 'slowing growth' by failing to lend Treasury cash to small businesses. |
| 17 | <i>The Times</i> | How savers lose a 'hidden' £19bn a year. |
| 18 | <i>The Times</i> | Co-operative Bank set to fail latest stress test. |
| 19 | <i>The Times</i> | Fears over Chinese slowdown after trade data disappoints. |
| 20 | <i>The Telegraph</i> | Six triggers that could renew a eurozone crisis in 2015. |

Spanish texts, Level 1

| Text ID | Source | Title |
|---------|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | <i>El País</i> | Bruselas advierte de que la banca española sigue siendo vulnerable. <i>Brussels warns that Spanish banking remains vulnerable.</i> |
| 2 | <i>El Mundo</i> | Francia recortará 21.000 millones en 2015 y no cumplirá con el déficit. <i>France will cut 21 billion in 2015 and will not meet the deficit.</i> |
| 3 | <i>El Mundo</i> | El Santander gana un 26% menos tras provisionar casi 3.000 millones. <i>Santander gained 26% less after provisioning almost 3,000 million.</i> |
| 4 | <i>El Mundo</i> | Barclays, Santander, Unicaja y La Caixa son los bancos con las comisiones más elevadas. <i>Barclays, Santander, Unicaja and La Caixa are the banks with the highest commissions.</i> |
| 5 | ABC | El déficit del Estado se situó en el 3,11% en septiembre, un 9% menos que hace un año. <i>Public deficit stood at 3.11% in September, 9% less than a year ago.</i> |
| 6 | <i>El País</i> | Los precios de la mitad de los bienes y servicios del IPC caen al cierre de 2014. <i>Prices of half of the goods and services of the CPI fall at the end of 2014.</i> |
| 7 | ABC | La economía española crece un 1,4% en 2014, su primera subida desde 2008. <i>Spanish economy grew by 1.4% in 2014, its first rise since 2008.</i> |
| 8 | <i>El País</i> | El BBVA eleva su previsión de crecimiento al 2,7% en 2015 y 2016. <i>BBVA raises its growth forecast to 2.7% in 2015 and 2016.</i> |
| 9 | <i>El País</i> | El Banco de España calcula que la recesión se frenó en el inicio de 2013. <i>Bank of Spain calculates that the recession slowed down in early 2013.</i> |
| 10 | <i>El Mundo</i> | La prima de riesgo nota el efecto 'Syriza-Podemos'. <i>Risk premium notes the 'Syriza-Podemos' effect.</i> |
| 11 | ABC | El manguerazo de Draghi deja la prima de riesgo en mínimos de 2010. <i>Draghi's 'manguerazo' leaves the risk premium at 2010 lows.</i> |
| 12 | <i>El Mundo</i> | Los bancos que no cumplan los requisitos de capital exigidos no podrán pagar bonus a sus directivos. <i>Banks that do not meet the required capital requirements will not be able to pay bonuses to their managers.</i> |

| Text ID | Source | Title |
|---------|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 13 | <i>El País</i> | Récord inversor de firmas de Estados Unidos en sociedades españolas. <i>Investing record of United States firms in Spanish companies.</i> |
| 14 | ABC | El Banco de España mejora ocho décimas, hasta el 2,8%, la previsión de crecimiento en 2015. <i>Bank of Spain improves eight tenths, to 2.8%, the growth forecast in 2015.</i> |
| 15 | ABC | El crédito al consumo crece un 8,25% y encadena ya dos años al alza. <i>Consumer credits grow 8.25% and has already been on the rise for two years.</i> |
| 16 | ABC | El plan de pago a proveedores impulsó el PIB entre 0,3 y 0,6 puntos. <i>Payment plan to suppliers pushed GDP between 0.3 and 0.6 points.</i> |
| 17 | <i>El País</i> | España cobra por primera vez por emitir deuda a corto plazo. <i>Spain charges for the first time for issuing short-term debt.</i> |
| 18 | <i>El Mundo</i> | Los bancos se niegan a reembolsar intereses negativos a sus clientes. <i>Banks refuse to reimburse negative interests to their customers.</i> |

English texts, Level 2

| Text ID | Source | Title |
|---------|------------------------|----------------------------------------------------------------------------------------|
| 1 | <i>World Finance</i> | Banking networks are key to unlocking trade in the Middle East. |
| 2 | <i>Financial Times</i> | ECB blow to European bank backstop. |
| 3 | <i>Financial Times</i> | Central bank acts to ease China cash crunch fears. |
| 4 | <i>World Finance</i> | Investors look to long-term European gains. |
| 5 | <i>The Economist</i> | How the world's fourth-most-populous country is weathering the emerging-market turmoil |
| 6 | <i>The Economist</i> | Risk on. |
| 7 | <i>The Economist</i> | Not for turning. |
| 8 | <i>The Economist</i> | London has more to lose than most when it comes to scaring off oligarchs. |
| 9 | <i>The Economist</i> | The chancellor hands more freedom to retirees. |
| 10 | <i>The Economist</i> | Lessons from behavioural economics can boost tax compliance. |
| 11 | <i>Financial Times</i> | Banks under pressure on UK ringfencing plans. |
| 12 | <i>Financial Times</i> | FCA forex probe to accuse banks of control failings. |
| 13 | <i>World Finance</i> | Diversify to succeed: Bank of the West promotes international investment. |
| 14 | <i>Financial Times</i> | Britain and the cuts: Blow for Cameron as UK faces deeper cuts. |
| 15 | <i>World Finance</i> | Can Ecuador's digital currency save its economy? |
| 16 | <i>World Finance</i> | Will copying the Dutch pension system help the UK? |
| 17 | <i>World Finance</i> | Investors look to long-term European gains. |
| 18 | <i>Financial Times</i> | UK financial watchdog to review soaring level of fines. |
| 19 | <i>Financial Times</i> | FCA toughens rules for loan brokers. |
| 20 | <i>The Economist</i> | Job not done. |

| Text ID | Source | Title |
|---------|---------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 19 | ABC | Banco Santander supera por primera vez los 100.000 millones de capitalización en Bolsa. <i>Banco Santander surpasses for the first time the 100,000 million of capitalization in Stock Exchange.</i> |
| 20 | El País | La gran banca rebaja por primera vez los activos tóxicos del ladrillo. <i>Big banks reduce for the first time the toxic assets of the brick.</i> |

Spanish texts, Level 2

| Text ID | Source | Title |
|---------|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Cinco Días | Los fondos más conservadores del mercado sufrirán pérdidas en 2015. <i>The most conservative funds in the market will suffer losses in 2015.</i> |
| 2 | Expansión | Suiza penaliza con un $-0,75\%$ los tipos de los depósitos y 'libera' al franco. <i>Switzerland penalizes with -0.75% deposit rates and 'free' the franc.</i> |
| 3 | Expansión | La banca española supera la nuevas exigencias de capital del BCE. <i>Spanish banking surpasses the ECB's new capital requirements.</i> |
| 4 | Expansión | Sólo el 17% del Ibex informa de su contribución tributaria total. <i>Only 17% of the Ibex reports its total tax contribution.</i> |
| 5 | Banca15 | Caixabank espera crédito positivo de la cartera sana en 2015. <i>Caixabank expects positive credit from the healthy portfolio in 2015.</i> |
| 6 | Finanzas.com | Acciones e índices ganadores tras el 'QE' expandido de Mario Draghi. <i>Winning actions and indexes after the expanded 'QE' of Mario Draghi.</i> |
| 7 | Finanzas.com | El BCE comprará 100.000 millones de deuda pública española. <i>ECB will buy 100 billion Spanish public debt.</i> |
| 8 | Cinco Días | Alemania rompe previsiones con un crecimiento trimestral del 0,7%. <i>Germany breaks forecasts with quarterly growth of 0.7%.</i> |
| 9 | Cinco Días | Los precios cayeron un 1,3% en enero. <i>Prices fell 1.3% in January.</i> |
| 10 | Banca15 | El beneficio del Santander se situó en 5.800 millones en 2014. <i>Santander's profit stood at 5.800 million in 2014.</i> |
| 11 | Expansión | El acuerdo sobre el reparto de las indemnizaciones se alcanzó al aceptar Bankia ser el primero en pagar. <i>The agreement on repartition of compensation was reached by accepting Bankia to be the first to pay.</i> |
| 12 | Expansión | Banco de España: el BCE debe exigir provisiones suficientes a todos los bancos. <i>Bank of Spain: The ECB must demand adequate provisions from all banks.</i> |

| Text ID | Source | Title |
|---------|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 13 | <i>Expansión</i> | Draghi garantiza nanciación a coste de derribo para los Gobiernos a décadas vista. <i>Draghi guarantees funding at the cost of demolition for decades.</i> |
| 14 | <i>Cinco Días</i> | El BCE prepara una 'prima de riesgo' permanente y estable. <i>ECB prepares a 'permanent and stable' risk premium.</i> |
| 15 | <i>Banca15</i> | La falta de extraordinarios reduce el beneficio de la banca hasta marzo. <i>Lack of extraordinary reduces the benefit of the bank until March.</i> |
| 16 | <i>Finanzas.com</i> | La CNMV ya busca gestora y depositaria para los fondos de Banco Madrid. <i>CNMV is already looking for manager and depository for the funds of Banco Madrid.</i> |
| 17 | <i>Finanzas.com</i> | Alfombra roja para las rentas medias-altas. <i>Red carpet for middle-high incomes.</i> |
| 18 | <i>Cinco Días</i> | España entra en el club de la deuda a tipos negativos. <i>Spain enters the debt club at negative rates.</i> |
| 19 | <i>Banca15</i> | La banca española pone el foco en recuperar la rentabilidad perdida. <i>Spanish banks put the focus on recovering lost profitability.</i> |
| 20 | <i>Finanzas.com</i> | La UE alerta: a más deuda, más ajuste fiscal en España. <i>EU warns: to more debt, more fiscal adjustment in Spain.</i> |

References

- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: OUP.
- Anderson, Lloyd B. 1986. Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In *Evidentiality: The Linguistic Coding of Epistemology*. Wallace Chafe & Johanna Nichols (eds.), 273–312. Norwood, NJ: Ablex.
- Berdasco-Gancedo, Yolanda. 2016. *Análisis del grado de especialización de los textos de carácter económico: estudio contrastivo entre inglés y español*. PhD dissertation, Complutense University of Madrid.
- Biber, Douglas. 1993. Representativeness in Corpus Design. *Literary and Linguistic Computing* 8(4): 243–257. doi:10.1093/lc/8.4.243
- Boye, Kasper. 2012. *Epistemic Meaning: A Crosslinguistic and Functional-cognitive Study*. Berlin: De Gruyter. doi:10.1515/9783110219036
- Boye, Kasper & Harder, Peter. 2009. Evidentiality: Linguistic categories and grammaticalization. *Functions of Language* 16(1): 9–43. doi:10.1075/fol.16.1.03boy
- Cabré, M. Teresa. 1999. *Terminology: Theory, Methods, and Applications* [Terminology and Lexicography Research and Practice 1]. Amsterdam: John Benjamins. doi:10.1075/trlp.1
- Carretero, Marta. 2016. Noun phrases as expressions of evidentiality: An analysis of four English abstract nouns and their Spanish equivalents. *Kalbotyra* 69: 29–58.

- Celle, Agnès. 2009. Hearsay adverbs and modality. In *Modality in English: Theory and description*, Raphael Salkie, Pierre Bussutil & Johan van der Auwera (eds), 268–293. Berlin: De Gruyter. doi:10.1515/9783110213331.269
- Chojnicka, Joanna. 2012. Reportive evidentiality and reported speech: Is there a boundary? Evidence of the Latvian oblique. In *Multiple Perspectives in Linguistic Research on Baltic Languages*, Aurelija Usonienė, Nicole Nau & Ineta Dabašinskienė (eds), 170–192. Newcastle upon Tyne: Cambridge Scholars.
- Coates, Jennifer. 1983. *The Semantics of the Modal Auxiliaries*. London: Croom Helm.
- de Haan, Ferdinand. 2005. Encoding speaker Perspective. In *Evidentials: Linguistic Diversity and Language Theories* [Studies in Language Companion Series 72], Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds), 379–397. Amsterdam: John Benjamins. doi:10.1075/slcs.72.18haa
- Faller, Martina. 2002. *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. PhD dissertation, Stanford University.
- Fernández del Moral, Javier & Francisco Esteve Ramírez. 1996. *Fundamentos de la información periodística especializada*. Madrid: Síntesis.
- Fox, Renata. 1999. The social identity of management ergolect”. *English for Specific Purposes* 18: 261–278. doi:10.1016/S0889-4906(97)00057-4
- Givón, Talmy. 1984. *Syntax: A Functional-Typological Introduction*. Amsterdam: John Benjamins. doi:10.1075/z.17
- Hennemann, Anja. 2012. The epistemic and evidential use of Spanish modal adverbs and verbs of cognitive attitude. *Folia Lingüística* 46(1): 1–39.
- Hoffmann, Lothar. 1998. *Els llenguatges d'especialitat: Selecció de textos*. Barcelona: Universitat Pompeu Fabra.
- Huddleston, Rodney. 1995. The case against a future tense in English. *Studies in Language* 19(2): 399–446. doi:10.1075/sl.19.2.04hud
- Lampert, Günther. 2015. Sense activation triggering in English epistentials: Attention distribution, contextual modulation of meaning, and categorization issues. *Cognitive Semantics* 1: 77–103. doi:10.1163/23526416-00101003
- Larreya, Paul. 1984. *Le possible et le nécessaire. Modalités et auxiliaires modaux en anglais britannique*. Poitiers: Nathan Recherche.
- Lavid, Julia, Marta Carretero & Juan Rafael Zamorano-Mansilla. 2016. A linguistically-motivated annotation model of modality in English and Spanish: Insights from MULTINOT. *Linguistic Issues in Language Technology* 14(4): 1–33.
- Lavid, Julia, Marta Carretero & Zamorano Mansilla, Juan Rafael. 2017. Epistemicity in English and Spanish: An annotation proposal. In *Evidentiality and Modality in European Languages: Discourse-pragmatic Perspectives*, Juana I. Marín-Arrese, Julia Lavid-López, Marta Carretero, Elena Domínguez Romero, M. Victoria Martín de la Rosa & María Pérez Blanco (eds), 241–276. Bern: Peter Lang.
- Li, Charles N. 1986. Direct speech and indirect speech: A functional study. In *Direct and Indirect Speech*, Florian Coulmas (ed.), 29–45. Berlin: Mouton de Gruyter. doi:10.1515/9783110871968.29
- Lichtenberk, Frantisek. 1995. Apprehensional epistemics. In *Modality in Grammar and Discourse* [Typological Studies in Language 32], Joan Bybee & Suzanne Fleischman, 293–327. Amsterdam: John Benjamins. doi:10.1075/tsl.32.12lic
- Lyons, John. 1977. *Semantics*, Vol. 2. Cambridge: CUP.

- Mañoso Pacheco, Lidia. 2017. *Reporting Expressions: An English-Spanish Contrastive Study of Quality Press*. PhD dissertation, Complutense University of Madrid.
- Marín-Arrese, Juana I. 2011. Effective vs. epistemic stance and subjectivity in political discourse: Legitimising strategies and mystification of responsibility. In *Critical Discourse Studies in Context and Cognition* [Discourse Approaches to Politics, Society and Culture 43], Christopher Hart (ed.), 193–223. Amsterdam: John Benjamins. doi:10.1075/dapsac.43.10mar
- Marín-Arrese, Juana I. 2013. Stancetaking and inter-subjectivity in the Iraq Inquiry. Blair vs. Brown. In *English Modality: Core, Periphery and Evidentiality*, Juana I. Marín-Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds), 411–445. Berlin: Mouton de Gruyter. doi:10.1515/9783110286328.411
- Martínez-Egido, José J. 2015. Modalidades de enunciado y de enunciación en la transmisión del discurso económico y financiero en la prensa española: La estructura de la información. In *El español lengua de la comunicación profesional*, Andreu van Hooft, Anna Escofet, Kristi Jauregi, Bob de Jonge, Paula Lorente & Lieve Vangehuchten (eds), 7–18. Spanish Ministry of Education, Culture & Sport; Education Council of Belgium, Netherlands and Luxembourg.
- Nuyts, Jan. 2001. *Epistemic Modality, Language and Conceptualization: A Cognitive-pragmatic Perspective* [Human Cognitive Processing 5]. Amsterdam: John Benjamins. doi:10.1075/hcp.5
- Ruskan, Anna. 2015. Evidential adverbs in Lithuanian: A corpus-based study. *Kalbotyra* 67: 104–130. doi:10.15388/Klbt.2015.8945
- Searle, John R. 1976. A classification of illocutionary acts. *Language in Society* 5(1): 1–23. doi:10.1017/S0047404500006837
- Simon-Vandenberg, Anne-Marie & Aijmer, Karin. 2007. *The Semantic Field of Modal Certainty: A Corpus-based Study of English Adverbs*. Berlin: De Gruyter.
- Smirnova, Alla V. 2009. Reported speech as an element of argumentative newspaper discourse. *Discourse & Communication* 3: 79–103. doi:10.1177/1750481308098765
- Usonienė, Aurelia & Šinkūnienė, Jolanta. 2013. A cross-linguistic look at the multi-functionality of the English verb *seem*. In *English Modality: Core, Periphery and Evidentiality*, Juana I. Marín-Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds), 281–316. Berlin: Mouton de Gruyter. doi:10.1515/9783110286328.281
- Vatrican, Axelle. 2015. Evidentiality and epistemic modality in the rumor/journalistic conditional in Spanish. In *Evidentiality and the Semantics-Pragmatics Interface*, Bert Cornillie & Juana I. Marín-Arrese (eds). *Belgian Journal of Linguistics* 29: 83–100.
- White, Peter R. R. 2012. Exploring the axiological workings of ‘reported voice’ news stories – Attribution and attitudinal positioning. *Discourse, Context & Media* 1: 57–67. doi:10.1016/j.dcm.2012.10.004
- Wiemer, Björn & Stathi, Katerina. 2010. The database of evidential markers in European languages. A bird’s eye view of the conception of the database (the template and the problems hidden beneath it). *STUF* 63(4): 275–289.

Evidentiality and epistemic modality in Old Catalan

A diachronic cognitive approach to the semantics of modal verbs

Andreu Sentí

Universitat de València

Epistemic modality and evidentiality are two categories that have not been clearly defined in the literature. In order to clarify the boundaries between them, I draw a detailed semantic map for Catalan modal verbs *deure* ('must'), *haver de* ('have to') and *poder* ('can/may') in the Old period (11th–16th centuries). On the one hand, the modals *deure* and *haver de* develop an evidential reading (inferential process). On the other hand, an epistemic possibility value arises in the case of *poder*, which is not based on any explicit premise. I show that all these verbs are subjectively construed in different degrees and argue that a clear distinction between the subjective values and an evidential or epistemic interpretation is paramount.

Keywords: evidentiality, inference, epistemic modality, possibility, dynamic modality, subjectification

1. Evidentiality and epistemic modality: State of the art

In this chapter I shed some light on the boundaries between inferential evidentiality and epistemic modality. I claim that they need to be considered as two separated functional categories based on the analysis of diachronic data from Romance modal verbs *deure* 'must', *haver de* 'have to' and *poder* 'can/may', in particular from Old Catalan (11th–16th centuries).

In previous corpus-based studies I have studied in detail the grammaticalization and subjectification of Catalan modal verbs *deure* 'must', *haver de* 'have to' and *poder* 'can/may' (Sentí 2015a, 2015b, 2016, 2017), within the framework of Cognitive Grammar (Langacker 1987, 1991, 2006; Pelyvás 2000, 2006). In this

chapter I analyze all these verbs together in order to get a deeper understanding of their degree of subjectification as well as to provide a better account of how modal and evidential values are distributed.

The lexical verbs *deure* 'owe' and *haver* 'have, possess' first developed as auxiliaries expressing necessity, like their counterparts in other languages. After that, they acquired new nuances, whether inferential or epistemic. These modal constructions thus underwent a semantic change and the emergence of innovative values took place. Cognitive Grammar provides a theoretical background appropriate to account for these modal constructions. The cognitive approach is based on force dynamics (Talmy 1988) and conceptualizes the necessity reading of such verbs in the following terms: the **imposer** (the authority) forces the **doer** (the subject) to accomplish a purposeful action. In some cases, though, the imposer and the force (the conceptual origin of the necessity) are gradually fading away, uncovering a subjective conceptualization of the scene. Thus, the conceptual scene changes and new meanings arise in diachrony. In this chapter I analyse the subjectification process of all these verbs, specially focusing on the relation between modality and evidentiality.¹

The corpus used is the Old Catalan Corpus (CICA), a representative corpus of Old Catalan with typologically different texts.

1.1 Evidentiality and epistemic modality

Epistemic modality and evidentiality are two categories that have not been clearly defined in the literature. Numerous connections between the two notions make it difficult to properly establish the limits of each concept, especially in the case of Germanic and Romance languages (Chafe 1986; Palmer 1986; van der Auwera & Plungian 1998; Squartini 2016).

Epistemic modality indicates the degree of commitment of the speaker to the truth of the sentence; that is to say, the degree of probability of the state of affairs. Some modal adverbs as *probablement* 'probably', *possiblement* 'possibly' or *potser* 'maybe' and also modal verbs such as *poder* 'may' or *deure* 'must', as well as the Romance future tense are seen as having an epistemic meaning:

- (1) *Joan probablement vindrà demà.*
'John will probably come tomorrow'

1. See Sentí (2015a, 2015b) for an in-depth cognitive account of *haver de* and *deure*, and their subjectification process, based on Cognitive Grammar (Langacker 2006; Pelyvás 2000, 2006).

- (2) *Potser la xiqueta té fam.*
‘Maybe the girl is hungry’
- (3) *Agafa un paraigua, que pot ploure.*
‘Take an umbrella, because it may rain’
- (4) *Gerard deu ser a casa.*
‘Gerard must be at home’
- (5) *On és Rosa? No ho sé, serà a casa.*
‘Where is Rosa? I’m not sure, she will be at home’

On the other hand, evidentiality encodes the source of information. An evidential marker indicates the way the speaker has acquired information. In his traditional classification of evidential meanings, Willett (1988) distinguished two types: direct evidentials (i.e. the speaker has attested the facts) and indirect evidentials (i.e. the speaker has not witnessed personally the facts). The indirect source is split into two subcategories: reported information (hearsay markers), as in (6), or inferred information by the speaker based on some piece of evidence or on her/his own reasoning (inferentials), as in (7) and (8):

- (6) *Es veu que Joan vindrà demà.*
‘It seems that John will come tomorrow’
- (7) *Sembla que hi ha algú ací.*
‘It seems that there is someone here.’
- (8) *El riu va ple, deu haver plogut tota la nit.*
‘The river is full, it must have been raining all night.’

As can be appreciated from the previous examples, the relation between modality and evidentiality is close, and it is thus not surprising that several authors have linked the two concepts (Dendale 1994; Dendale & De Mulder 1996; Floyd 1999; Nuyts 2001; Cornillie 2007). The connection between the two values may seem quite obvious. On the one hand, epistemic modality is inevitably based on some piece of evidence that allows the prediction of the probability. On the other hand, evidential values have also been related to an epistemic meaning. In this sense, recall that direct evidential markers have been linked to epistemic certainty, and indirect evidential markers to lower degrees of certainty (Frajzyngier 1985; Willett 1988; Bybee, Perkins & Pagliuca 1994: 180; Nuyts 2006: 11).

However, the link between modality and evidentiality has been called into question in several works (de Haan 1999; Squartini 2004; Cornillie 2009; Sentí & Cornillie in press). In the next section I will deal with this issue in the light of modal verbs in Romance.

1.2 Evidentiality and modal verbs in Romance

The case of the semantics of *deure* ‘must’, *semblar* (‘seem’), *poder*² ‘can/may’ and their Romance and Germanic counterparts³ has remained a strongly controversial issue in the literature. These verbs have been considered either epistemic or evidential (inference subtype). The link (or confusion) between both values is actually quite striking in the particular case of (semi)auxiliary verbs in Germanic and Romance languages: recall the epistemic example of *deure* ‘must’ provided in (4), and the one as an inferential evidential marker in (8). Example (4) could be understood as epistemic because there is not any inferential process behind the auxiliary *deure*. However, in Example (8), there is an inferential background: because the river is full, it can be deduced that it must have been raining all night.

The relation between epistemic modality and inferential evidentiality has been approached in the literature from three different perspectives: (a) the traditional perspective argues that these verbs are modal verbs with an epistemic interpretation (Le Querler 2001; Vettters 2004; Gavarró & Laca 2002; Schmid 2012; Aijmer 1997; de Haan 2001); (b) another view has considered French *devoir* as a mixed evidential/modal marker (Kronning 1996, 2003); (c) a third perspective has argued that ‘must’ and its counterparts are evidentials in their primary reading, whereas the probability reading is a secondary one, derived from the inferential process (Dendale 1994; Dendale & Tasmowski 2001; Dendale & De Mulder 1996; Floyd 1999; Nuyts 2001; Cornillie 2007).

In the third approach, direct evidentiality is linked to epistemic certainty, and indirect evidential markers to lower degrees of certainty (Frajzyngier 1985; Bybee et al. 1994: 180; Nuyts 2006: 11). However, the link between evidential and epistemic meanings has been the subject of discussion in several works (de Haan 1999; Dendale & Tasmowski 2001; Nuyts 2006). Actually, there are data showing that several evidentials from different languages do not have such a direct relation with particular modal values (de Haan 1999; Squartini 2004; Cornillie 2009; Sentí & Cornillie in press). Therefore, although in some languages (e.g., Italian, Spanish, Catalan)⁴ both notions may converge in the same markers, this does not imply that they constitute one and the same category (see de Haan 1999 for Western Tarahumara, English and Dutch).

2. *Poder* ‘can/may’ is a polysemous verb. As explained in § 2.4 *poder* expresses dynamic modality, deontic modality, epistemic modality, permission, and other meanings.

3. This is the case of verbs *deure* (Cat), *deber* (Sp, Por), *devoir* (Fr), *dovere* (It), and their counterparts *must* (English) and *moeten* (Dutch), or *poder* (Cat, Sp, Por), *pouvoir* (Fr), *potere* (It) (‘may’).

4. On evidential markers in Contemporary Catalan see González (2011, 2014).

Before dealing with the relation between modality and evidentiality and analyzing the Catalan data, it would be useful to explain the typology of inferential evidentiality. There are two different inferential values attested in several typological studies (Anderson 1986; Aikhenvald 2004; Squartini 2008). Following Squartini's terminology, I differentiate between specific and generic inferences. In the first subtype, specific inferences are based on an external, directly perceived piece of evidence. In the second subtype, i.e. generic inference, the inferential reasoning is not based on observed, external evidence, but relies on previous knowledge of the speaker instead. Finally, I will differentiate these two meanings from pure conjectures, in which the speaker makes reference neither to external evidence nor to general world knowledge.

These typological meanings have been attested in several languages such as Tuyuca, Wintu, Tsafiki and Tariana (Squartini 2008: 923), as well as in Romance. Squartini (2008: 922–924) thus shows that the French verb *devoir* and the Italian verb *dovere* can express specific (9) and generic inferences (10), but no conjectures (11), which are expressed by means of an epistemic future (12) (cf. Tasmowski 2001):

- (9) [*Indicano un ragno*] *Attento, deve essere ancora vivo, perché ho visto che si muove*
 '[Pointing to a spider] Be careful! It **must** still **be** alive, for I saw it moving'
- (10) [*Suonano alla porta*] *Deve essere il postino*
 '[The doorbell rings] It **must be** the postman'
- (11) ??[*Suonano alla porta*] *Non aspettavo nessuno; deve essere Gianni.*
 '[The doorbell rings] I was not expecting anybody. It **must be** Gianni'
- (12) [*Suonano alla porta*] *Non aspettavo nessuno; sarà Gianni.*
 '[The doorbell rings] I was not expecting anybody. It **might be** Gianni'

As mentioned before, it has often been claimed that the type of inference made by the speaker suggests a *natural* association with the speaker's commitment to the truth of the state of affairs. For instance, when the speaker makes a specific inference, the associated epistemic commitment is usually considered strong. Yet, the question is whether inferences can always be associated with epistemic values (cf. Dendale 1994; Squartini 2008; Cornillie 2009).

In this chapter, this question is addressed. My claim is that there is no direct semantic correlation between inferencing and epistemic modality. In line with Cornillie (2009), I stress that the type of inference does not correspond to a specific type of epistemic commitment, as I show using diachronic data for the Catalan modal verbs *deure* 'must' and *haver de* 'have to'. In addition, I argue that the function of the conjecture value is to express the speaker's commitment to the situation, as opposed to the source of information. To support this claim I use diachronic data for the Catalan modal verb *poder* 'can/may'.

2. An inferential reading for *deure*, *haver de* and *poder* in Old Catalan?

The grammaticalization of Catalan modal verbs from Vulgar Latin to the 16th century entails important semasiological changes. The three verbs – *deure* (‘must’), *haver de* (‘have to’) and *poder* (‘can/may’) – developed several epistemic or evidential values from dynamic or deontic values, especially in the 15th and 16th centuries (Sentí 2015a, 2015b, 2016, 2017; cf. Yllera 1980; Garachana 2017; Bybee et al. 1994).

In what follows, I deal with the verbs *deure* and *haver de*, arguing that they are inferential evidentials in Old Catalan (§ 2.1, § 2.2, § 2.3). After that, I will contrast these data with *poder*, which I consider to be an epistemic marker (§ 2.4).

2.1 The modal verb *deure*: Specific vs. generic inferences

The verb *deure* developed inferential values in the 13th century, but it was not until the second half of 14th century that the frequency of these uses become more significant (Sentí 2015a, 2015b). In the 16th century, this value become quite widespread. It represents 30% of tokens as displayed in Table 1.

Table 1. Core values of the modal *deure* in Old Catalan

| | 11th–1250 | 1250–1300 | 1300–1350 | 1350–1400 | 15th | 16th |
|---------------|-----------|-----------|-----------|-----------|--------|--------|
| Necessity | 85.96% | 84.72% | 92.97% | 75.68% | 72.06% | 61.58% |
| Evidentiality | 0.88% | 0.8% | 1.62% | 6.9% | 7.87% | 29.94% |
| Future | 6.14% | 11.02% | 2.7% | 13.22% | 16.52% | 3.95% |

The oldest evidential values that *deure* developed correspond to a *specific inference*. As shown in (13) and (14), there are specific pieces of evidence that form the objective basis from which the speaker infers the state of affairs. For example, in (14) the speaker knows that the ring must belong to the knight Curial, because there is a lion printed on it and Curial always stamps his letters with this lion.⁵ I use boldface for epistemic expressions and the modal verb; whereas the underlining corresponds to the external evidence:

- (13) *Hauida entre ells gran discepció sobre açò, concordaren ensemps que, per tal com jo havia experimentat cascuna natura, fos jutge de la qüestió dessus dita, axí com aquell qui mils ho devia saber que altri.* (Metge, Somni: 190, 1398 [ms. XV])
 ‘They had a long discussion, and they finally agreed that, since I had experience of all kinds, I should be the judge in that above-mentioned question, because I was the one who **had to know** better than anyone.’

5. Cf. Torró (2012: 273) for an interpretation of *feya leó per amor de la Güelfa* (Curial: 312).

- (14) *e donaren-li lo jupó e l'anell que havien comprat; e Melchior féu juyhí que aquell anell degué ésser estat de Curial, per rahó del leó que tenia tallat, car tots temps Curial feya leó per amor de la Güelfa.*

(*Curial*: 312, 1435–1445 [ms. 1445–1458])

‘and they gave him the jerkin and the rig that they had bought; and Melchior **thought** that that ring **must have belonged** to Curial, because of the lion that it had sculpted, because Curial always stamps his letters with this lion.’

Importantly, in the 15th century, some constructions developed another evidential value, the generic inference (Squartini 2008; Sentí 2015a). The verb *deure* now conveys an inference that is not based on the observation of external facts, but on the previous knowledge of the speaker. The construction has lost its objective component and, hence, has a bigger subjective import: there is a stronger link between the speaker and the semantic value of the modal verb (Sentí 2015a). In the 16th century, the majority of tokens with evidential meaning express a *generic inference* (more subjective), as in the following examples:

- (15) *Miraren tots a Aznar, e feren rahó que aquest devia ésser valent cavaller e molt forts, encara que fos molt jove e tendre*

(*Curial*: 217, 1435–1445 [ms. 1445–1458])

‘They all looked at Aznar, and **came to the conclusion** that he **must be** the brave and strong knight, even though he was very young’

- (16) *e tenint gran desig de poder-ho manifestar a son cosín germà, Diafebus, duch de Macedònia, stimant que, axí com ell, a cascuna persona devia plaure la contentació que ell tenia.*

(*Tirant*: 1020, 1460 [ms. 1490])

‘and having a great wish to tell that to his cousin and brother Diafebus duke of Macedonia, estimating that, like him, every person **must have liked** the happiness that he was feeling.’

- (17) *Lo emperador, que sabé que lo conestable era allí, pensà Tirant hi devia ésser tanbé.*

(*Tirant*: 855, 1460 [ms. 1490])

‘The emperor, who knew that the general was there, thought that Tirant **must have been** there too’.

- (18) *Margarit crec que s'aturarà en Catalunya y que-s deu penedir de les despeses fetes fins así.*

(*Estefania*: carta 89, 1522–1542)

‘I **think** that Margarit will stop in Catalonia and he **must regret** the expenses made until today’.

- (19) *Pense que deu ser altre ab qui pensava enviar-la.*

(*Estefania*: carta 96, 1522–1542)

‘I **think** it **must be** someone else who he wanted to send it with.’

- (20) Don Pedro: Què lletania tan llarga! Com hi **deuen estar** instruïts!

(*Col·loquis*: 188, 1557)

‘Don Pedro: What a long litany! How instructed **must they have been!**’

2.2 The modal verb *haver de* ‘have to’: Specific inference

Unlike the verb *deure*, the verbal periphrasis *haver de* + infinitive did not develop widespread inferential values in Old Catalan (Sentí 2017). The inferential reading emerges only from the 16th century on, and its presence is reduced to the 6% of the tokens as shown in Table 2.

Table 2. Unambiguous values of *haver de* + infinitive (‘have to’)

| | 12th–14th | 15th | 16th |
|--------------------------------|-----------|--------|--------|
| Participant-external necessity | 71.18% | 55.55% | 35.04% |
| Deontic necessity | 3.49% | 17.68% | 25.55% |
| Future | 18.78% | 24.24% | 32.12% |
| Evidential inferential | | | 6.57% |

The inferential value developed by *haver de* is the specific inference:

- (21) *Deffensen a Déu provant que, com a just, no castiga a nengú sens culpa, y que, si nengú éss castigat per Déu, necessàriament à de ser injust y mal home*
 ‘They defend God by showing that, as he is fair, he doesn’t punish anyone who is not guilty, and that if someone is punished by God, he must necessarily be a bad and unfair man’
(Grandeses: 37, XVIa)
- (22) *Y té dit circo de larch ·CLX· canas de Monpaller, abans més que manco, perquè los càrcers avien de estar ont estan are las casas. Que jo no é mesurat sinó sols lo que are és encara plaça, y de ample té de cornisa a cornisa ·coranta...*
 ‘And this circus is 160 *canas de Montpeller* [a measure unit] long, more or less, because the prisons **had to be** where the houses are now. I have only measured out the part that today is still the square...’
(Grandeses: 180, XVIa)
- (23) *yo tinch per cert que vostra opinió acerca de las novenas és la millor y que lo que diu Carbonell és desvarió y, com haveu dit, temeritats y res saber, y par-me que vós haveu de ser molt cregut en assò, puix no és adnumerat vostre llinatge en les novenes y perquè no-s porà dir que passió de propri interés vos fa parlar així.*
(Col·loquis: 183, 1557)
 ‘I know for sure that your opinion about the *novenas* is the best one and that what Carbonell says is non-sense and, as you have said, is nothing but bullshit, and **it seems to me that you have to be** very experienced in this matter, because your name is not mentioned in the *novenas*, and thus it won’t be possible to say that you speak in your own interest’.

For instance, in (21), the speaker says *Déu no castiga a nengú sens culpa* ‘God doesn’t punish anyone who is not guilty’, therefore, *si nengú éss castigat per Déu, necessàriament à de ser injust y mal home* ‘if someone is punished by God he **must necessarily**

be a bad and unfair man' (see underlining). In short, *haver de* has developed only the first stage of subjectivity of inferential values, that is to say, specific inferencing.

2.3 *Deure* and *haver de*: Subjectification, evidentiality and epistemic modality

The diachronic data presented show that the evidential values developed by modal verbs are progressively more and more subjective. The evidential value that emerges from the deontic modality, both in the case of *deure* and *haver de*, is the specific inference, which is more objective. But the verb *deure*, as it grammaticalized, developed a more subjective value, i.e. generic inference, since the conceptualizer is included in the scope of predication.⁶ Thus, taking into account the corpus data presented here, a more fine-grained semantic path of the verbs can be described as a tendency towards more subjective meanings:

NECESSITY > SPECIFIC INFERENCE > GENERIC INFERENCE

In the course of its grammaticalization process, the verb *deure* developed a more subjective value, the generic inference (Langacker 2006; Pelyvás 2000, 2006). Evidence of subjectification is the fact that the inferential uses of *deure* are often accompanied by hedges such as the verbs *parèixer* ('to seem'), *conèixer* ('to know'), *pensar* ('to think'), *estimar* ('to calculate'), *creure* ('to believe'), *fer juí* ('to judge'), as in (15) to (20) - highlighted in bold. These collocations are especially found in the generic inference, the more subjective one; and even in some, less subjective, specific inferences. On the other side *haver de*, which expresses a specific inference, only co-occurs with the form *par* ('it seems') in a single token, in (23). Thus, *haver de* behaves like French *devoir* and German *müssen* and *sollen* which normally do not co-occur with subjective hedges, as shown by Dendale (1994); and Mortelmans (2000, 2002). At this point we can see very clearly that Old Catalan *deure* is more subjectified than its French and German counterparts. Therefore, I postulate that the difference between *deure* and *haver de* is that the former has become construed in a more subjective way.

Thus, at the end of the period (16th century), the semantic distribution was the following as indicated in Table 3.

6. The evidential meaning of *haver de* and *deure* arises from deontic modality in the grammaticalization process. Deontic conceptualization is construed in an objective view: the authority forces the doer to carry out the situation. Conceptualization is then subjectified in the new inferential meanings, especially in generic inference, and the original force is fading away. Cf. Sentí (2012, 2015a) for a cognitive account of grammaticalization of *haver de* and *deure*, based on Cognitive Grammar (Langacker 2006; Pelyvás 2000, 2006).

Table 3. Values of *haver de* and *deure* in the 16th century

| | <i>Haver de</i> | <i>Deure</i> |
|--------------------|-----------------|--------------|
| Necessity | ✓ | ✓ |
| Specific inference | ✓ | ✓ |
| Generic inference | × | ✓ |

Another relevant issue is the relation between evidentiality and epistemic modality. In this study, the meanings described above are considered evidentials. It is true that sometimes a specific inferential value can be related to a derived epistemic meaning of certainty (cf. § 2.1, Dendale 1994). In fact, in our corpus, there are some examples of *deure* that encode an evidential meaning, and they might also be related to a degree of certainty, as can be observed in (24) to (27). For instance, in (25), the speaker deduces that the knight ‘must not have felt safe’ because his opponent was strong and muscular. This is a case of specific inference from which a certainty meaning can be derived. However, what contributes to the epistemic reading is not exactly the verb *deure*, but rather other lexical elements that co-occur with the modal verb, such as *sens tota falla* (‘with no doubt’), *sens dubte* (‘without a doubt’), *tinc per molt cert* (‘I am quite certain’):

- (24) *Lo cor sentit de Tirant, prestant presumí que algun cars havia seguit a la princessa, segons los grans crits que ell sentia donar a hòmens e a dones, e cregué fermament que axí devia ésser.* (Tirant: 749, 1460 [ms. 1490])
 ‘Tirant’s heart immediately felt that something had occurred to the princess, judging from the screams of women and men that he heard, and he **firmly believed** that this **must be** the case.’
- (25) *e de grans espatles, e fort espés de tots los membres, e axí era tan fort, que sens tota falla lo cavaller qui ab ell combatia no-s devia tenir per segur*
 ‘and he had big shoulders and had a very muscled body, and he was so strong that **with no doubt** the knight that was fighting him **must not be** free of fear’ (Curial: 107, 1435–1445 [ms. 1445–1458])
- (26) *Ayxí que totas estas cosas donan testimoni molt sert del que devia ésser esta ciutat de Tarragona, que gosave de la gran felicitat y glòria romana...*
 (Grandeses: 205, XVIa)
 ‘So all these things give evidence very truly of what the city of Tarragona **must be**, and of the fact that the city enjoyed a great Roman happiness and glory’...
- (27) *Don Pedro: Segons això, per cert tinch que degueren restar poch militars assí en la conquesta com tant los extrengueren.* (Col·loquis: 124, 1557)
 ‘Don Pedro: According to that, I **know for sure** that they **must be** very few soldiers left since during the conquest many of them died.’

Crucially, we also find generic inference uses combined with lexical elements referring to certainty:

- (28) *Per cert tinch que deu haver-hi alguna cosa natural y secreta o serà per algun contàgio. Teniu-ne vosaltres de assò ohit res a vostres antinchs?*

(Col-loquis: 200, 1557)

‘I know for sure that it must be something natural and secret, or it may be due to some contagion. Have you heard anything about that?’

- (29) *Per cert tinch yo que en Cathalunya deu haver-hi molts llocs que no seran per als propis senyors de tanta utilitat.*

(Col-loquis: 194, 1557)

‘I know for sure that in Catalonia there must be lots of places that won’t be so useful for the lords’.

In (29) the speaker says *deu haver-hi molts llocs* ‘there must be a lot of places’, but without basing his conclusion on either any explicit premise or any external evidence. This is thus an example of generic inference that should be linked to a lower degree of probability. However, in this case the speaker also expresses certainty when he says *Per cert tinch yo* ‘I know for sure’. Therefore, I believe that the possible epistemic meanings (certainty) that could be seen in these examples actually do not derive from the inference type of the modal verbs. The epistemic commitment is an additional meaning derived from other expressions (*Per cert tinch yo*). Consequently, two different dimensions should be distinguished, inferential evidentiality and epistemic modality.

2.4 The modal verb *poder* ‘can/may’: Epistemic modality

The modal *poder* prototypically expresses dynamic possibility, which refers to an ability of the subject of the clause (Palmer 1979). Following Nuyts (2006), I will distinguish **participant-inherent dynamic** (capacities completely inherent to the first participant: *Anna pot córrer* ‘Anna can run’); **participant-imposed dynamic** (potentials determined by local circumstances that cannot be controlled by the participant: *La porta és oberta, ja pots passar* ‘The door is open, you can enter’); and **situational dynamic** (potentials inherent to the whole situation: *A l’hivern, pot nevar* ‘In winter it can snow’).

According to our corpus data,⁷ the dynamic possibility value represents around 70% of tokens during the period studied – see the first row in Table 4. In some cases, the modal verb has a high degree of vagueness between several values (dynamic,

7. Because of the high degree of frequency of the modal verb *poder*, I analyzed the 7.7% of all tokens in the selected texts from the CICA corpus (Senti 2016).

deontic, epistemic) – see the top of the table⁸. In Old Catalan, other meanings also arise, such as deontic and permission, and an epistemic value.⁹

Epistemic possibility is the peripheral semantic value that most clearly developed in Old Catalan. In the 15th century, it represents the 8.3% of the tokens, whereas in the 16th century it consolidates with a 14.4% proportion.

Table 4. Values of *poder* in Old Catalan

| | 1100–1250 | 1250–1300 | 14th | 15th | 16th |
|----------------------|-----------|-----------|-------|------|-------|
| Dynamic | 94.1% | 73.8% | 72.8% | 67% | 69.3% |
| Dynamic / epistemic | | 9.1% | 5.9% | 4.8% | 9.2% |
| Dynamic / permission | | 4.8% | | 3.1% | 0.7% |
| Dynamic / deontic | | | 0.7% | 0.4% | 0.7% |
| Dynamic /future | | | 2.2% | 1.3% | |
| Right | 5.9% | 0.5% | 0.7% | | |
| Future | | 1% | 2.2% | 0.9% | 0.7% |
| Future / permission | | | 1.4% | | |
| Future / deontic | | 0.5% | | | |
| Deontic | | 1.6% | 2.9% | 4.4% | 2% |
| Deontic / permission | | 0.5% | 0.7% | | 0.7% |
| Permission | | 1.6% | 6.6% | 7.9% | 2.6% |
| Epistemic | | 5.8% | 3.7% | 8.3% | 14.4% |
| Surprise | | 0.5% | | 0.4% | |
| Concessive | | | | 0.4% | |
| Desire | | | | 0.4% | |
| Complaint | | | | 0.4% | |

I now turn to an analysis of the innovative tokens that express an epistemic possibility meaning in order to examine the relation between epistemic modality and *deure* and *haver de*, and provide a better understanding of the epistemic and evidential paradigms.

The epistemic possibility meaning arises from situational dynamic modality. Both of them express close meanings. Moreover, they usually have inanimate, generic or impersonal subjects, and they usually combine with stative infinitives. Situational dynamic modality expresses a possibility inherent to the state of affairs. On the other hand, in the case of epistemic modality possibility is an evaluation by the conceptualizer.¹⁰ In some tokens, both meanings can be understood, like in the following examples:

8. The ambiguous (or transitional) meanings are separated with a backslash in the table.

9. Cf. Sentí 2016 for a corpus-based diachronic account of the Catalan modal verb *poder*.

10. In Nuyt's (2006: 6) terms, epistemic modality can be described as "the estimation of the chances that the state of affairs expressed in the clause applies in the world".

- (30) *açò és cosa que-ns pesa molt, e podets-ho conèixer, que axí só vengut jo entre vós e que en vós me fiy e en vostra amor e que us hé en cor* (Fets: 36, 13th C.)
 ‘this is something important for us, and you **can/may know** it, because I came in front of you and I trust you and I trust your love and I carry you in my heart’
- (31) *Interrogata si aquesta testis viu la dita Anthònia ab lo dit Francí o ab altre sola de què pogués ésser sospita contra la dita dona.* (Marquès: 63; 14th)
 ‘She was asked whether she, a witness, saw Anthònia alone with Francí or with someone else, for which reason there **could be** suspicion against that woman,’

For example, (31) can be understood as (a) as situational dynamic possibility (‘Anthònia was seen with Francí or someone else that has the attributes to be suspicious’); or (b) an epistemic possibility (‘Anthònia was seen with Francí or someone else and the conceptualizer hypothesizes that maybe they were suspicious’).

Although there are tokens with an unambiguous epistemic meaning from the 13th century onwards, the epistemic *poder* consolidated in the 15th and especially the 16th century. The epistemic value that arises from *poder* is pure conjecture.¹¹ The conceptualizer bases his hypothesis or conjecture only on his own belief. Therefore, *poder* specifically developed a subjective value, as in (32) to (38):

- (32) *E ·I· dia, mentra que Nonnós, baron de reverència, pensàs que aquel loc pogre ésser bon a nodrir ·I· poc d’erbolam per condiment, si aquela roca no-l tengués ocupat* (Diàlegs: 25, 13th)
 ‘One day, while Nonnós, respected baron, thought that that place **could/might** be good to nourish some herbs if that rock did not keep him busy.’
- (33) *no és mal ans és molt bé y cosa necessària qu’escrigua també la defensa, puix pot ser bona y seria falta molt gran, -al parer meu- que de escriure se dexàs.* (Col·loquis: 54, 16th)
 ‘It is very necessary that you also write the defence, since it **can/may be** good and in my opinion it would be a huge mistake not to write it’
- (34) *Jo é sentida la pena que vostra senyoria pot pensar, però estic bona y Lloyset bonísim* (Estefania: carta 16, 16th)
 ‘I have felt the pity that you **can/may think** about, but I feel good now and Lluïset too.’
- (35) *e continent e, si és en Xàtiva, procureu promptament de envyar-la-y perquè importa y, en la tarda, poria haver inconvenyent.* (Estefania: carta 153; 16th)
 ‘You should sent it there quite early, because it is important and in the afternoon there **might be** some inconvenient.’
- (36) *axí perquè les noves que li puc dir de nostra salut són bones y també poria ser que nostre Senyor nos agés fet la mercè que tots desijam* (Estefania: carta 8, 16th)
 ‘The news I can give you about my health are good and it **might also be** that the Lord had given us the mercy that we all want’

11. See Squartini (2008) for the differences between inferences and pure conjecture.

- (37) *perquè poria ser que lo dit Pelliser se detingés més del que diuen.*

(Estefania: carta 64, 16th)

‘because it **might be** that the mentioned Pelliser stopped more than they say.’

- (38) *tostemps que-l veiga li diré mon parer y no descrech que li à pesat tot lo que s'à fet. Porà ser que ell matex mourà abans de molt lo que à refusat.*

(Estefania: carta 4; 16th)

‘When I see him I will tell him my opinion and I believe that he is affected by all what happened to him. It **will be possible** that he will start acting very soon.’

For instance, in (32), the conceptualizer thought that that place could nourish well some herbs, therefore he ‘believed’ that it might be a good place. There is no trace of any external source that allows for the inference. Therefore, the important element is not the source of information, but the hypothesis and the fact that the speaker is downgrading his commitment to the truth of the statement.

Unlike *deure* and *haver de*, the verb *poder* does not show any inferential process or reasoning; there is no external evidence. I argue that the basic meaning of *poder* is a conjecture in which the source of information is not relevant, and what is highlighted is epistemic commitment. This hypothetical meaning is stressed when the modal verb *poder* appears in the conditional or future tense, as in (35)–(38). For example, in (37), the possibility that ‘Pelliser stopped more than they say’ is a hypothesis suggested by the speaker, and it is not inferred from any premise. The epistemic commitment of the speaker is reduced because of the conditional tense.

Although some tokens could be seen as going against my claim, I will show that they are not real counterexamples in the next section.

2.4.1 Possible counterexamples?

In very few tokens, one could see an evidential background upon which the speaker bases himself in order to make an inference. In (39) and (40), there is external evidence in the context (see the underlined words). For instance, in (39) the speaker says ‘with the remains and the ruins that we can see, everyone can understand that the city could actually be like that’. We could think that that is a specific inference.

- (39) *Jo hé llegit en un libre molt vell de lletra de mà, que per la vellesa sua y anti-
quitat és digne de crèdit, que era la ciutat de Tarragona ans de la sua roïna de
sexanta_ quatre_ milia. Ab los vestigis y gran roïna que vuy se mostra, vent-la
pot cada qual comprendre que podia ésser així. (Grandeses: 103; 16th)*

‘I have read in a very ancient book (that truly deserves my trust) that Tarragona used to have 64.000 inhabitants. Now, with the remains and the ruins that we can see, everyone can understand that the city **could/might** actually **be** like that.’

- (40) *De assò no puch donar sertitut ninguna que fos ciutat; pot ésser ayxí tant solament, que los vells y antichs ho tenen y diuven per aver-ho hoit a dir a sos avis y predecessós, los quals també ho tenien y avien hoit dels antichs.*

(Grandeses: 107, 16th)

‘I cannot claim for sure that it was a city; it **can/may be** like that only, **because the old men think so and say** that they have heard this from their ancestors, who in turn had also heard from their ancestors.’

However, even in all these cases I think that the main function of the epistemic verb *poder* is to indicate conjecture, and not the source of information (inference). I defend this hypothesis for three reasons.

In the first place, I think that the evidential background in (39) and (40) is not due to a reasoning process, or to an inferential process. These examples actually express a meaning semantically closer to situational dynamic modality, more objectively construed. I believe that this proximity to the original dynamic meaning is what makes us think of a more objective value. Epistemic possibility is based on a real possibility inherent to the very same situation, as in the case of ambiguous Examples (30) and (31) seen above. Therefore, *poder* is not an evidential marker that encodes the source of information, i.e. inference.

In the second place, we saw that in most cases there are no explicit premises in the context, unlike what we saw with *deure* and *haver de*. Examples (32)–(38) express epistemic possibility and there is no trace of any inferential process. The conceptualizer uses *poder* to suggest a hypothesis.

And finally, in the third place, the epistemic verb *poder* can conceptualize facts that the speaker considers to be possible in the reality of the past, the present, and crucially also the future (see Examples from (35) to (38)). On the contrary, the inferential verbs *deure* and *haver de* can only conceptualize the past or the present because in this case the speaker conceptualizes reality from the premises that s/he knows or had known. Actually, in Old Catalan, the auxiliary *deure* in an evidential reading cannot appear in the conditional or the future (Sentí 2017: § 4.4.6). Possibly, this is due to the inferential process marked by the modal verb *deure* the conceptualizer observes something (in the present or in the past) and this piece of evidence allows him/her to infer some information that is necessarily linked to a state of affairs in the present or the past – crucially, *deure* as an evidential marker cannot express future events. However, in the case of *poder*, there is no inferential process, therefore the conceptualizer can express the possibility of present, past and future states of affairs.

In short, in the conceptualization of reality, *poder* signals that the speaker makes a hypothesis, and the verb does not encode the source of information, it rather downgrades epistemic commitment.

3. Conclusion

In this chapter, I have shown that a corpus-based approach is useful to describe the evidentiality and modality semantic categories. After a comparative analysis of *deure*, *haver de* and *poder* in diachrony (from 11th to 16th centuries), I have shown how their grammaticalization process led to the rise of subjective values. The modals *deure* and *haver de* developed inferential evidential meanings from deontic and external necessity. In particular, the Catalan modal *haver de* developed a specific inference, not really subjectified and closer to necessity. Regarding the modal *deure* it is grammaticalized in a deeper way and developed specific, but also generic inference, relying on more subjective conceptualization. Finally, unlike *deure* and *haver de*, the modal verb *poder* give rise to a value of conjecture, an epistemic modal value. Therefore, the three modal verbs become subjectified in different ways, but their subjective construal should not be confused with epistemic modality and degrees of certainty.

A diachronic account of modal verbs sheds light on how new values arise. In some cases, an inferential process can be observed: the context includes external sources from which the speaker infers information, as in the case of *deure* and *haver de*. However, in other cases, as with the verb *poder*, it is hard to identify an inferential process.

In the grammaticalization process of these modal verbs, there is a tendency towards subjectification, in which modal force is progressively attenuating (Sentí 2015a, 2015b). But this subjectification process should not be confused with a higher or lower degree of certainty, that is to say epistemic meanings. The results presented here show that there is not necessarily a connection between an evidential value and an epistemic one. Instead, a careful examination of modal and evidential values allows us to distinguish one category from the other, as has been proposed elsewhere (de Haan 1999; Squartini 2008; Cornillie 2009). In the case of Old Catalan, each auxiliary specializes in a particular novel meaning around the 16th century: specific inference (*haver de*), generic inference (*deure*), and epistemic possibility (*poder*).

In Contemporary Catalan, the semantics of modal verbs has evolved. Future research should focus on how the construction with *deure* evolves into an even more subjective conceptual schema, given the existence of evidence for that (for example, *deure* can be used in interrogative sentences). Likewise, an analysis of the epistemic future will certainly be crucial in order to provide a better understanding of the modal and evidential categories as a whole.

References

- Aijmer, Karin. 1997. 'I think' and English modal particle. In *Modality in Germanic Languages*, Toril Swan & Olaf Jansen Westvik (eds), 1–47. Berlin: Mouton de Gruyter. doi:10.1515/9783110889932.1
- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: OUP.
- Anderson, Lloyd B. 1986. Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In *Evidentiality: The Linguistic Coding of Epistemology*, Wallace Chafe & Johanna Nichols (eds), 273–312. Norwood NJ: Ablex.
- Bybee, Joan L., Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago IL: University of Chicago Press.
- Chafe, Wallace. 1986. Evidentiality in English conversation and academic writing. In *Evidentiality: The Linguistic Coding of Epistemology*, Wallace Chafe & Johanna Nichols (eds), 261–272. Norwood NJ: Ablex.
- CICA = *Corpus Informatitzat del Català Antic (CICA)*, J. Torruella (dir.), Manuel Pérez Saldanya, Josep Martines & Vicent Martines (text selection). <http://www.cica.cat> [query 2009–2016]
- Cornillie, Bert. 2007. *Evidentiality and Epistemic Modality in Spanish (Semi-)Auxiliaries: A Cognitive-functional Approach*. Berlin: Mouton de Gruyter.
- Cornillie, Bert. 2009. Evidentiality and epistemic modality. On the close relationship between two different categories. *Functions of Language* 16(1): 44–62.
- de Haan, Ferdinand. 1999. Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics* 18: 83–101.
- de Haan, Ferdinand. 2001. The relation between modality and evidentiality. *Linguistische Berichte* 9: 201–216.
- Dendale, Patrick & De Mulder, Walter. 1996. Déduction ou abduction: Le cas de devoir inférentiel. In *l'Énonciation médiatisée*, Zlatka Guentchéva (ed.), 305–318. Louvain: Peeters.
- Garachana Camarero, Mar (ed.). 2017. *La gramática en la diacronía: la evolución de las perífrasis verbales modales en español*. Madrid: Iberoamericana; Frankfurt: Vervuert.
- Dendale, Patrick & Tasmowski, Liliane. 2001. Introduction: Evidentiality and related notions. *Journal of Pragmatics* 33: 339–348. doi:10.1016/S0378-2166(00)00005-9
- Dendale, Patrick. 1994. *Devoir épistémique, marqueur modal ou évidentiel? Langue Française* 102: 24–40. doi:10.3406/lfr.1994.5712
- Floyd, Rick. 1999. *The Structure of Evidential Categories in Wanka Quechua*. Dallas TX: The Summer Institute of Linguistics.
- Frajzyngier, Zygmunt. 1985. Truth and the Indicative Sentence. *Studies in Language* 9: 243–254. doi:10.1075/sl.9.2.05fra
- Gavarró, Anna & Laca, Brenda. 2002. Les perífrasis temporals, aspectuals i modals. In *Gramàtica del català contemporani*, Vol. 3, Joan Solà, Maria Rosa Lloret, Juan Mascaró & Manuel Pérez Saldanya (eds), 2663–2726. Barcelona: Empúries.
- González, Montserrat. 2014. Evidentiality, intersubjectivity and salience in Spanish and Catalan markers *claro/clar* and *la verdad/veritat*. Special Issue *Evidentiality in Discourse*. *Intercultural Pragmatics* 11(3): 411–436. doi:10.1515/ip-2014-0019
- González, Montserrat. 2011. Indirect evidence in Catalan. A case study. In *The Pragmatics of Catalan*, Lluís Payrató & Josep M. Cots (eds), 145–172. Berlin: Mouton de Gruyter. doi:10.1515/9783110238693.145

- Kronning, H. 1996. *Modalité, cognition et polysémie: Sémantique du verbe modal 'devoir'*. Uppsala & Stockholm: Almqvist & Wiksell International.
- Kronning, H. 2003. Modalité et évidentialité. In *Aspects de la modalité*, Merete Birkelund, Gerhard Boysen & Poul Søren Kjaersgaard (eds), 131–151. Tübingen: Max Niemeyer. doi:10.1515/9783110963434.131
- Langacker, Ronald Wayne, 1987. *Foundations of Cognitive Linguistics, Vol. I: Theoretical Prerequisites*. Stanford CA: Stanford University Press.
- Langacker, Ronald Wayne. 1991. *Foundations of Cognitive Linguistics. Vol. II: Descriptive Application*. Stanford CA: Stanford University Press.
- Langacker, Ronald Wayne. 2006. Subjectification, grammaticization, and conceptual archetypes. In *Subjectification: Various Paths to Subjectivity*, Angeliki Athanasiadou, Costas Canakis & Bert Cornillie (eds), 17–40. Berlin: Mouton de Gruyter. doi:10.1515/9783110892970.17
- Le Querler, Nicole. 2001. La place du verbe modal pouvoir dans une typologie des modalités. In *Les verbes modaux*, Patrick Dendale & Johan van der Auwera (eds), 17–32. Amsterdam: Rodopi.
- Mortelmans, Tanja. 2000. On the 'evidential nature of the 'epistemic' use of the German modals *müssen* and *sollen*. *Special issue Modal Verbs in Germanic and Romance Languages*, Johan van der Auwera & Patrick Dendale (eds). *Belgian Journal of Linguistics* 14: 131–148.
- Mortelmans, Tanja 2002. 'Wieso sollte ich dich küssen, du hässlicher Mensch!' A study of the German modals *Sollen* and *Müssen* as grounding predications in interrogatives. In *Grounding: The Epistemic Footing of Deixis and Reference*, Frank Brisard (eds), 391–432. Berlin: Mouton de Gruyter. doi:10.1515/9783110899801.391
- Nuyts, Jan. 2001. Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics* 33: 383–400. doi:10.1016/S0378-2166(00)00009-6
- Nuyts, Jan. 2006. Modality: Overview and linguistic issues. In *The Expression of Modality*, William Frawley (ed.), 1–26. Berlin: Mouton de Gruyter.
- Palmer, Frank Robert. 1979. *Modality and the English Modals*. London: Longman.
- Palmer, Frank Robert. 1986. *Mood and Modality*. Cambridge: CUP.
- Pelyvás, Péter. 2000. Metaphorical extension of 'may' and 'must' into the epistemic domain. In *Metaphor and Metonymy at the Crossroads*, Antonio Barcelona (eds), 233–250. Berlin: Mouton de Gruyter.
- Pelyvás, Péter. 2006. Subjectification in (expressions of) epistemic modality and the development of the grounding predication. In *Subjectification: Various Paths to Subjectivity*, Angeliki Athanasiadou, Costas Canakis & Bert Cornillie (eds), 121–150. Berlin: Mouton de Gruyter. doi:10.1515/9783110892970.121
- Schmid, Beatrice. 2012. Un aspecto sintáctico del *Curial e Güelfa*: Las perífrasis modales. In *Novel·la cavalleresca anònima del segle XV en llengua catalana. Linguistic and Cultural Studies on 'Curial e Güelfa', a 15th Century Anonymous Chivalric Romance in Catalan* [IVITRA Research in Linguistics and Literature 3], Antoni Ferrando (ed.), 843–874. Amsterdam: John Benjamins.
- Sentí, Andreu & Cornillie, Bert. In press. The rise of the evidential readings of the Catalan periphrasis *deure* + infinitive. In *From Composite Predicates to Verbal Periphrases in Romance Languages*, Mar Garachana, Sandra Montserrat & Claus Pusch (eds). Amsterdam: John Benjamins.
- Sentí, Andreu. 2015a. Subjectification and attenuation in the conceptual schema of the Catalan modal verb *deure* with evidential meaning. *eHumanista/IVITRA* 8: 517–543.

- Sentí, Andreu. 2015b. Modal verbs, future and grammaticalization in Old Catalan. A cognitive approach. *Catalan Journal of Linguistics* 14: 179–198.
- Sentí, Andreu. 2016. Modality and evidentiality in Old Catalan. Grammaticalization and subjectification of the modal verb *poder* ‘can/may’. Talk given at 49 Annual Meeting of the Societas Linguistica Europaea, Naples.
- Sentí, Andreu. 2017. Modalitat i evidencialitat en català antic. *Un acostament cognitiu a les perífrasis verbals amb deure i amb haver*. Barcelona: Publicacions de l’Abadia de Montserrat / Institut Interuniversitari de Filologia Valenciana.
- Squartini, Mario. 2004. Disentangling evidentiality and epistemic modality in romance. *Lingua* 114(7): 873–889. doi:10.1016/S0024-3841(03)00064-0
- Squartini, Mario. 2008. Lexical vs. grammatical evidentiality in French and Italian. *Linguistics* 46(5): 917–947. doi:10.1515/LING.2008.030
- Squartini, Mario. 2016. Interaction between modality and other semantic categories. In *The Oxford Handbook of Modality and Mood*, Jan Nuyts & Johan van der Auwera (eds), 50–67. Oxford: OUP.
- Talmy, Leonard. 1988. Force dynamics in language and cognition. *Cognitive Science* 2: 49–100. doi:10.1207/s15516709cog1201_2
- Tasmowski, Liliane. 2001. Questions au conditionnel. In *Le conditionnel en français*, Patrick Dendale & Liliane Tasmowski (eds), 331–343. Paris: Klincksieck.
- Torró, Jaume. 2012. El manuscrit del *Curial e Güelfa*. Els pròlegs i el filòcola. *RLM XXIV*: 296–281.
- van der Auwera, Johan & Plungian, Vladimir A. 1998. Modality’s semantic map. *Linguistic Typology* 2: 79–124. doi:10.1515/lity.1998.2.1.79
- Vetters, Carl. 2004. Les verbes modaux pouvoir et devoir en français. *Revue Belge de Philologie et d’Histoire* 82: 657–671. doi:10.3406/rbph.2004.4851
- Willett, Thomas. 1988. A cross-linguistic survey of the grammaticalization of evidentiality. *Studies in Language* 12: 51–97. doi:10.1075/sl.12.1.04wil
- Yllera, Alicia. 1980. *Sintaxis histórica del verbo español: Las perífrasis medievales*. Zaragoza: Universidad de Zaragoza.

‘I think’

An enunciative and corpus-based perspective

Graham Ranger

Université d’Avignon

This chapter focuses on the sequence ‘I think’ as a discourse marker, used in evidential or epistemic contexts. ‘I think’ is seen to assume a variety of different values, which Kaltenböck (2010), among others, identifies as “shielding”, “approximator”, “structural” or “booster” functions. I hypothesise that ‘I think’ is not inherently ambiguous, but that different values reflect specific configurations, which depend on identifiable contextual features. The present study explores this hypothesis, first with a corpus-based investigation of collocational affinities of the sequence, which reveals a number of characteristic environments. Secondly, I elaborate an enunciative description of ‘I think’ in terms of a basic schematic form, which undergoes certain controlled and calculable deformations to generate local “shapes” (Culioli 1990). I conclude that ‘I think’ in itself expresses neither evidentiality nor epistemic modality, but that these result from specific contextual configurations.

Keywords: evidentiality, epistemic modality, Theory of Enunciative Operations, schematic form, corpus linguistics

1. Introduction

Parentetical ‘I think’, as illustrated by Examples (1)–(3), entertains links with both evidentiality and modality.

- | | |
|--------------------------------------------------------------------------------------|-----------------------|
| (1) <i>I think</i> people should adopt an attitude of live and let live! | FL5 397 |
| (2) Rupert does, <i>I think</i> [pause] tend to get through a hell of a lot of milk. | KBL 3213 |
| (3) I dunno, it was either Clive or Richard <i>I think</i> . | KE6 5293 ¹ |

1. Examples are taken from the British National Corpus, accessed *via* the BNCweb interface and identified by alphanumeric text identifier and sentence references. Available online at <<http://bncweb.lancs.ac.uk/>> Cf. also Hoffmann et al. (2008).

It is related to evidentiality in so far as it represents a “mode of knowing” or “source of information”.² It is additionally related to epistemic modality – serving in some contexts to reduce speaker commitment, as in (3), and in others to reinforce commitment, in (1) and arguably in (2) (Aijmer 1997).

The present corpus study aims to link on the one hand, the approach of enunciative linguistics – and more specifically the Theory of Enunciative and Predicative Operations (henceforth TEPO) – with its fine-grained attention to context and to formal modelisation, and, on the other, certain types of quantitative data, collocations in particular.

I will begin by presenting a limited number of key issues that emerge from previous studies on ‘I think’. Next, I will conduct a corpus-based investigation of certain collocational and positional affinities of the sequence. Then I will endeavour to show how the statistical data brought to light might tie in with an enunciative characterisation of ‘I think’. This will in turn lead to a final discussion on the evidential or epistemic character of ‘I think’ and on the representation of situated meaning in the TEPO.

2. Previous research

There is a vast amount of recent and insightful research on ‘I think’ all of which I cannot mention or do full justice to within this chapter.³ One might nonetheless identify three key points or problems addressed in various ways by different authors, relating, respectively, to the source of parenthetical ‘I think’, to its meanings and to those factors which contribute to selecting one meaning in preference to another.

2.1 Where does ‘I think’ come from?

Urmson’s (1952) *Parenthetical Verbs* is probably the first work to draw attention to the particular behaviour of certain verbs that form small matrix-clauses which accept initial, medial and final positions. He draws attention to differences in meaning between these parenthetical items and the usual lexical meaning of the verbs from which they derive. Urmson includes in the class verbs such as ‘regret, rejoice’, ‘deduce, infer’ or ‘know, believe’. Although he does not actually mention ‘I think’,

2. Chafe & Nichols (1986), Aikhenvald (2004).

3. In particular, for reasons of space and relevance, I will not here consider the full-length diachronic study by Palander-Collin (1999) or more recent work in a Construction Grammar framework by Van Bogaert (2010, 2011).

he does make parenthetical use of it within the article. Interestingly in the present context, Urmson (1952: 495–496) concludes that:

[Parenthetical verbs] themselves have not, in such a use, any descriptive sense but rather function as signals guiding the hearer to a proper appreciation of the statement in its context, social, logical, or evidential. [...] They help the understanding and assessment of what is said rather than being a part of what is said.

Urmson, whose reflexion needs to be situated in the tradition of the ordinary language philosophy of Austin or Strawson, for example, pays little attention to differences in the position of the parenthetical clause, or to the presence or absence of a that-complementiser. These questions are dealt with in the generative framework *via* a transformation known as *slifting* (Ross 1973), involving the deletion of the complementiser 'that' and the subsequent "S-lifting" of the object clause into main clause position. Thompson and Mulac (1991: 317) place this in a diachronic framework, with the hypothesis that epistemic parentheticals derive from a historical process of that-deletion and reanalysis:

[...] evidence suggests that the most frequent subjects and verbs occurring with what syntacticians have considered to be 'that-less' 'complements' [...] have in fact been reanalyzed by speakers as epistemic phrases, which have a degree of freedom not possible for subject-verb combinations.

Brinton (1996, 2008) has criticized what she has referred to as the "matrix-clause hypothesis", arguing against the purported diachronic movement from that-complements to that-less complements. The hypothesis nonetheless represents a widely accepted and influential explanation for the derivation of comment clauses in general and hence of 'I think' in particular.⁴

2.2 What does 'I think' mean?

A number of differentiable values or meanings of 'I think' – in many approaches referred to as functions – have emerged in the various studies devoted to the marker. For Lakoff (1975), initial 'I think' is included among certain hedges which are supposed to characterize women's language, as opposed to men's. Hedges such as 'sorta' might be used to "blunt the force of a rather painful assertion" (ibid: 79), but also in other cases "as an apology for making an assertion at all". She extends this to 'I think': "Another manifestation of the same thing is the use of 'I guess' and 'I think' prefacing declarations or 'I wonder' prefacing questions, which themselves are hedges on the speech-acts of saying and asking" (ibid: 79).

4. Aijmer (1997) also takes issue with this derivation of parentheticals.

Holmes (1990) takes issue with this undifferentiated analysis of the meaning of ‘I think’, pointing out the importance of syntactic position and intonation. She distinguishes one broad category labelled “tentative”, including expressions of uncertainty and softeners, and another labelled “deliberative”, which is on the contrary a sign of authority and which, for Holmes, is actually prevalent in women’s speech (Holmes 1990: 199–200).

A form’s lexical shape alone does not provide sufficient information to identify its function. Hence, like ‘I think’ and the tag question, ‘you know’ may be used either as a hedge or as an intensifier (or booster) [...] An analysis of women’s and men’s usage which treated all instances as hedges would clearly be unhelpful.

In a contrastive study based in part on English–Swedish translation preferences, Aijmer (1997) retains the two categories “deliberative” and “tentative”, adding a vaguer, “discourse function”, in particular when turn-initial ‘I think’ is clustered with other DMs: but ‘I think’, well ‘I think’, etc.

Simon-Vandenberg (2000) focuses on ‘I think’ in the genre of political interviews, compared with casual conversation. She again works with the two categories “deliberative” and “tentative”, but notes another, more rhetorical, use, in keeping with the discourse genre, when medial ‘I think’ marks a boundary between thematic and rhematic elements as in (2) above or in the following example in (4):

- (4) You must have been amazed yourself ‘I think’ by the scale and scope and vitality of Chinese society.” (Simon-Vandenberg 2000: 50)

More recently, Kaltenböck (2010: 257) proposes four categories for ‘I think’ according to the function it fulfils: (1) a “shield” – indicating a lack of commitment pragmatically associated with politeness or mitigation –; (2) an “approximator” – indicating semantic imprecision; (3) a “structural or filler function” – used in contexts of disfluency and / or topicalisation strategies; (4) a “booster” – reinforcing speaker commitment.

2.3 How are multiple meanings derived?

Most previous studies of ‘I think’ are not content merely to list different functions of ‘I think’, but seek to show how meanings might be disambiguated in context through various factors. At least five sets of factors have been recognised to play a role in determining how meanings are selected in context: position, scope, prosody, text type and linguistic context (or co-text). The position of ‘I think’ relative to its host – initial, medial or final, at least – has been claimed to correlate with specific meanings. Thus, for Aijmer (1997: 21), initial position, especially when followed by the complementizer ‘that’, yields “deliberative” ‘I think’, while medial and final positions necessarily give rise to “tentative” meanings.

Kaltenböck (2010) recognises the importance of position, but also points out the role of the *scope* of 'I think', showing that when clause-internal, phrasal elements are targeted, 'I think' typically operates as an "approximator" or hedge. In his spoken corpus, prosody contributes importantly to the identification of scope. Kaltenböck (2009a), for example, shows that 'I think' can be prosodically detached, left-bound, right-bound or both left- and right-bound. This is important in determining scope and from there, function.

Another factor of note is the situation or text type. Simon-Vandenberg (2000) shows that conversational uses of 'I think' prove frequently tentative, while in the political interview genre, the deliberative mode predominates. She concludes that "in political interviews, speakers do not primarily use 'I think' to express uncertainty but to convey 'this is my opinion'. The expression therefore suggests authority rather than hesitation" (ibid: 2000: 60).

Lastly, the strictly linguistic context, that is, the pressure from surrounding linguistic items often referred to as co-text, has received a certain amount of insightful comment (Aijmer 1997; Kaltenböck 2010; or Simon-Vandenberg 2000, in particular) but has not, to the best of my knowledge, been the object of systematic quantitative research. It is to this we now turn our attention.

3. Corpus-based investigation of collocational affinities

I will look at collocational affinities of 'I think' in the British National Corpus (BNC) accessed via the Phrases in English and BNCweb interfaces.⁵ Formal variants of 'I think' such as 'I do think', 'I would think' are excluded for reasons of space from the present study. 'I think that', however, is included, since it is considered to operate in much the same fashion as initial 'I think' without the complementizer 'that'.⁶

After some general considerations concerning frequencies by text type, genre and type of interaction, I move on to collocations. The first step here is to identify a number of characteristic clusters featuring 'I think'. These are then parsed for regularities in terms of parts of speech. The second step is to use these constructions to identify collocational affinities of 'I think' according to position and cluster features. The hope is that regularities observed in this way will provide clues to elaborating a semantico-syntactic profile for 'I think'.

5. The Phrases in English website is a front end for exploration of clusters or n-grams in the BNC developed by William H. Fletcher. URL: <http://phrasesinenglish.org/>

6. Kaltenböck (2010) chooses to exclude 'I think that' from the study, curiously, given that Kaltenböck (2009b) claims that there is no significant difference in prosody between initial 'I think' with and without 'that'. Kärkkäinen (2003) or Kaltenböck (2011) treat both forms together.

3.1 General considerations

In the entire BNC – both spoken and written sections – some 41,268 occurrences of the string ‘I think’ were found. Not all these occurrences can necessarily be qualified as the parenthetical discourse marker that interests us. In a random cross-genre sample of 500 occurrences, a small minority of 23 was categorized as lexical ‘I think’ while almost 92% were clearly parenthetical. The overall frequency for the string ‘I think’ is 419.76 pmw, which among similar comment clause sequences is a close second behind the sequence ‘you know’ (431.76 pmw).⁷ This normalized frequency leaps to 2481.69 pmw in the spoken section of the BNC compared to 175.58 pmw in the written part. Within the spoken BNC, the three genres in which ‘I think’ appears with the highest normalized frequencies are the oppositional contexts of public debate (4410.2 pmw), broadcast discussion (4032.33 pmw) and meetings (3894.46 pmw). In terms of interaction type, it is correspondingly more frequently attested in dialogue than in monologue.⁸

3.2 Significant n-grams featuring ‘I think’

In the next step, I use the Phrases in English website to check on how ‘I think’ forms characteristic clusters, specifically 5-grams, with other linguistic items. If one fixes a minimum threshold of a frequency of 20 occurrences for 5-grams, then this yields 107 5-gram types when ‘I think’ is initial in the cluster, that is, ‘I think’ * * * where the asterisk represents a wild card,⁹ 35 5-gram types for post-initial ‘I think’ (* ‘I think’ * * *), and only two 5-gram types for pre-final and final ‘I think’ (i.e. * * ‘I think’ * and * * * ‘I think’). These results reflect the fact that ‘I think’ is most frequent in clause initial position, and that there is a far greater degree of collocational affinity between ‘I think’ and what follows than between ‘I think’ and what comes before.

Let us now explore further this large set of 5-grams featuring initial ‘I think’. The Phrases in English website indicates the parts of speech involved in the

7. However, a random sample of 500 instances of the sequence ‘you know’ yielded 155 clearly lexical occurrences (31%), suggesting that the string ‘you know’ is far more likely to be lexical than the string ‘I think’. If so, then this would mean that ‘I think’ is the most frequent parenthetical discourse marker.

8. The figures quoted do not exclude potential non discourse marking instances, but since these are far and away the most frequent, the genre specific tendencies indicated here can be considered sufficiently representative.

9. Yielding, for example, 341 occurrences of “I think it’s a”, 229 occurrences of “I think it would be”, 131 of “I think that’s a”, and so on. Note that I differentiate between initial, medial or final relative to the cluster, or 5-gram, and clause-initial, etc.

n-gram results. On this basis it is possible to make further categorisation within the 107 5-grams obtained. In 88.7% of these 5-grams, 'I think' is followed by a personal pronoun, corresponding in most cases to an initial 'I think', without a that-complementiser. In over 85% of these cases the personal pronoun is followed by a form of the verb 'be' (49.5%), by a modal (26.1%) or by a form of the verb 'have' (10.2%). In summary, then, for cluster-initial 'I think' we obtain 'I think' + personal pronoun + BE (e.g. "I think it's..."), 'I think' + personal pronoun + HAVE (e.g. "I think I've...") and 'I think' + personal pronoun + MODAL (e.g. "I think he'll..."). Other significant configurations in 5-gram clusters with non-initial 'I think' include "I mean 'I think' *", and / "but I think * *" and "that's what / why I think".¹⁰

3.3 N-grams with cluster-initial 'I think'

Let us now focus in more closely on the three most common constructions with 'I think' in initial position, that is, 'I think' followed by a personal pronoun, followed in turn by be by have or by a modal.

In the first of these, of the 'I think' you ought to... type, the most salient modals, by log-likelihood, are: 'll, should, would, might, ought'. Particularly significant here is the presence of 'should', 'ought' and 'might', which figure far more prominently here than when the personal pronoun is *not* preceded by 'I think'. Table 1 illustrates this, with the second column indicating the overall frequency of modals after a personal pronoun, compared with the frequency of modals after the sequence 'I think' + personal pronoun in the third column.

Table 1. Modals ranked by Log Likelihood after a personal pronoun with and without 'I think' in a right-hand window of one word (BNC Spoken)¹¹

| Rank | _PNP + VM0 (default) | 'I think' _PNP + VM0 |
|------|----------------------|----------------------|
| 1 | 'll | 'll |
| 2 | can | should |
| 3 | could | would |
| 4 | 'd | might |
| 5 | would | ought |

10. Lohmann and Koops (2016) discuss just this type of discourse marker "sequencing", that is, possible combinations of discourse markers with other discourse markers.

11. The codes _PNP and _VM0 represent, respectively, any personal pronoun and any modal verb, in the BNCweb syntax.

On further investigation, the ‘I think’ personal pronoun + HAVE string also corresponds largely to the modal schema, too, since a majority of occurrences of ‘have’ here correspond to the semi-modal expressions ‘had better’ or ‘have got to’.

The most frequent configuration noted for initial ‘I think’ was ‘I think’ personal pronoun + BE. This is unsurprising given the high frequency of the personal pronoun + BE string in any case. More significantly, however, a search for adjectives within a two-word window after the string yields the results given in Table 2 (‘important’, ‘right’, ‘fair’, ‘good’, ‘better’...). It is interesting here to note that not one of the top ten adjectives ranked by log-likelihood in this sequence (third column) is among the top ten adjectives in the same sequence without ‘I think’ (second column). Intuitively it appears therefore to be the case that initial ‘I think’ primes the following context (or “co-text”) to favor subjectively loaded, evaluative adjectives, a point I shall be returning to later.¹²

Table 2. Adjectives ranked by Log Likelihood after a personal pronoun followed by the verb ‘be’ with and without ‘I think’ (BNC Spoken) in a right-hand window of 1–2 words¹³

| Rank | _PNP + VB + (default) | ‘I think’ _PNP + VB + |
|------|-----------------------|-----------------------|
| 1 | sure | important |
| 2 | afraid | right |
| 3 | sorry | fair |
| 4 | concerned | good |
| 5 | supposed | better |
| 6 | happy | true |
| 7 | glad | alright |
| 8 | aware | terrible |
| 9 | interested | disgusting |
| 10 | all | difficult |

3.4 N-grams with final ‘I think’

One way of testing the relevance of these results for initial ‘I think’ is to look at adjectives collocating this time with ‘I think’ in final position. Given the sometimes inconsistent transcription conventions for pauses, etc. in the BNC, it is not easy to find a satisfactory search string for final ‘I think’: I have therefore simply searched for typographical pauses, in the form of a comma or a full stop, after the sequence

12. “Priming” is used in the sense of Hoey (2005: 8) to refer to the way words and word sequences “become loaded with the contexts and co-texts in which they occur”.

13. The code _VB + represents any form of the verb BE.

'I think'.¹⁴ The results here, shown in Table 3, are very different from those noted when 'I think' is placed initially. In final position 'I think' remarkably collocates with a significantly high frequency of cardinals in the preceding context (eight of the first ten), that is, with adjectives that are very different from the evaluative type observed when 'I think' was initial.

Table 3. Adjectives ranked by Log Likelihood before final 'I think' in a left-hand window of 1–5 words (BNC Spoken)¹⁵

| Rank | _ADJ 'I think' [\,,\.] |
|------|------------------------|
| 1 | ninety |
| 2 | another |
| 3 | nineteen |
| 4 | fifty |
| 5 | important |
| 6 | four |
| 7 | eighteen |
| 8 | hundred |
| 9 | seventy |
| 10 | five |

The point of the explorations described in this section is to use targeted corpus enquiries to bring to light a number of significant constructions involving 'I think'. These include medial 'I think', in association with other discourse marking expressions ('and, but, well I think'), initial 'I think' followed by certain modal expressions (notably 'should', 'ought to', 'had better') or by certain evaluative adjectives ('important', 'right', 'fair') and final 'I think' preceded by certain numerical expressions. In the next section I seek to investigate how such significant correlations might help us to model a schematic form for 'I think'.

14. Such a search string will be high on *precision*, but not necessarily good on 'recall', since occurrences of 'I think' before pauses that are not indicated typographically will not be retained.

15. The search syntax indicates an adjective _ADJ, followed by 'I think', followed either by a comma or a full stop.

4. Modelisation within an enunciative perspective

4.1 The schematic form

The enunciative perspective on language argues in favor of an integrated pragmatics, claiming that many phenomena traditionally labelled ‘pragmatic’ can be explained as potentially part of the linguistic system itself, provided one recognises that linguistic items do not possess a fixed inventory of meanings, but mobilise abstract schematic forms which generate contextually situated shapes (Culioli 1990: 178). The schematic form of a marker is typically postulated in the course of fine-grained analysis of examples in context. The current study aims to show how corpus methodology, and collocations in particular, may provide quantitative elements to help explore this movement from abstract potential to situated values.

Earlier I noted at least four different “functions” of ‘I think’, *pace* Kaltenböck (2010), *viz*: “shield function” – tentative use, etc., “approximator function” – hedge, etc., “structural or filler function”, “booster function” – deliberative use, etc. Within an enunciative perspective, these “functions”, which correspond essentially to pragmatic phenomena, are considered to be no more than characteristic configurations of an abstract *schematic form*. The schematic form is defined by Culioli (1990: 116; my translation) in the following terms:

The analysis of a marker (understood not in the sense of a label, but as a **marker of an operation** or possibly of a polyoperation) must result in a formal representation possessing stable and verifiable properties. From this formal representation, which I call **schematic form**, emerge further forms which are in fact deformations of the basic form. The question is to understand how these deformable schemata are organized.

In the case of ‘I think’, I claim that ‘I think’ is not, in itself, a “shield”, a “booster”, etc. but rather that ‘I think’ in all configurations marks the localisation of a target notion *p* relative to a subjective representation of the speaker.¹⁶

Of course, the act of assertion always implies a form of subjective representation. The assertion of *p* involves: (1) commitment (wishing, meaning); (2) materialisation (saying, writing); (3) a subjective representation (thinking, believing, knowing); (4) a representation situated in a referential space, that is a “state of affairs” (Culioli 1999: 96). And so to assert “*p*” is for example: “to wish to say that one thinks (etc.) that <*p*> is the case”, (op. cit., translation mine).

16. The target notion in question corresponds frequently, but not always, to a proposition, particularly in certain occurrences of medial ‘I think’.

4.2 Knowing, believing, thinking

In the passage quoted, Culioli leaves open the question of the difference, crucial for our purposes, between 'thinking', 'believing', 'knowing', etc. These three predicates can be considered to differ in terms of subjective positioning, i.e. the way in which a speaker situates his/her representation of the notion p relative to other potential speakers, the speech community, etc.

- 'know', marks that a representation p is located relative to an enunciative source, and that this source is potentially identifiable with any other enunciative source. In other words, the speaker claims no particular subjective responsibility for the assertion of p .
- 'believe', on the other hand, marks that a representation p is located relative to an enunciative source, and that this source is potentially differentiable from other enunciative sources. In other words, the predicate makes explicit claims as to the subjectivity of the representation p , and anticipates other, potentially conflicting representations.
- 'think' like 'know' and 'believe', marks that a representation is located relative to an enunciative source, but, unlike 'know' and 'believe', 'think' leaves indeterminate the question of the position of the enunciative source "vis-à-vis" other enunciative sources. This will depend on contextual parameters, which can tip the balance in various directions.

The collocational affinities elucidated in Section 2, and the typical 'I think' constructions which these affinities enable us to identify can help us build hypotheses as to which contextual parameters appear most instrumental in disambiguating the values of 'I think'.

4.3 The position of 'I think'

The corpus searches made in Section 2 identified differences in construction types depending on the relative positions of 'I think' and the targeted notion p . 'I think' contributes to the determination of a target notion p and in this sense, 'I think' is a 'locator', as the term is used within the TEPO, while its target p , is a 'locatum'. In common with other locator / locatum relationships, the relative order plays an important part in determining the meaning.

When the locator is placed initially, the locator / locatum relationship is pre-constituted: the locator necessarily anticipates its locatum. When it is placed finally, the addition of the locator constructs the relationship retroactively, requiring us to readjust our perception of the locatum accordingly. In the case of 'I think', this

difference is often one between ‘I think’ *p*, framing a statement in terms of a speech act of subjective commitment from the outset, or on the contrary introducing “post hoc” some form of limitation to the preceding sequence – *p* ‘I think’.

- (5) And *I think* I should definitely be awarded my doctorate as soon as possible. AOD 2116
- (6) Yeah she’d been in there about [pause] seven or eight months *I think*. KCT 9089

Examples (5) and (6) are typical of these configurations. In (5), an initial ‘I think’ frames a proposition which itself contains a number of markers of subjective commitment – ‘should’, ‘definitely’ or ‘as soon as possible’. In (6), ‘I think’ follows a proposition containing a numerical indication, already mitigated by an adverb of approximation ‘about’ and a pause. The result, in (5), is a ‘deliberative’ (Holmes, Aijmer) or ‘booster’ (Holmes, Kaltenböck) effect and, in (6), a ‘hedging’, ‘tentative’ (Holmes) or ‘approximator’ (Kaltenböck) effect. One might argue that these effects are a result of the lexical input more than relative position. It is however not easy to separate the two factors, nor indeed desirable, since the point is that order and lexical choice are very closely interrelated.

4.4 Configurations with initial ‘I think’

It was noted in Section 2 that the most salient configurations for initial ‘I think’ involve certain modals, used deontically (‘should’, ‘ought to’) (7–8) or epistemically (*might*, ‘will’, etc.) (9–11) as well as certain evaluative expressions (‘*important*’, ‘*right*’, ‘*fair*’) (12–14), illustrated respectively below:

- (7) And *I think* we *ought* to get him over to one of our meetings. D97 142
- (8) *I think* you *should* study it very carefully and notice how you feel at the end of it. JYM 820
- (9) *I think*, when *I think* it since, *I think* he *must* have been a remarkable man to work and slave like that. G4R 50
- (10) I’ve done some now, *I think* they’ll be ready. KBB 4485
- (11) Well go, *I think* she’s *going* to be shattered, I’d better wake up Jo. KPY 1079
- (12) But there are targets, and and *I think* targets are *important* in order to shape a culture. J9D 592
- (13) *I think* it’s *disgusting*, they should do more for people. N2 327
- (14) Yes *I think* that’s *fair* enough. J9P 1049

What these expressions have in common is an assumed decision, the construction of a choice of alternatives, p versus $non-p$. This might be a deontic choice concerning courses of action, as in (7) and (8), an epistemic choice concerning possibility or probability, as in (9)–(11), or an evaluative choice bearing on rightness or wrongness, as in (12)–(14). In each case p is given priority by the speaker S_0 while $non-p$ is attributed to some other enunciative source in potential opposition, noted S^* .¹⁷

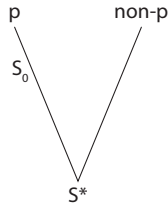


Figure 1. Differentiation between speaker endorsement S_0 of p and opposing positions S^*

In indicating the subjective nature of this assessment with ‘I think’, a speaker is potentially positioning him- or herself differentially relative to a co-speaker. In the majority of cases when ‘I think’ is placed initially, this differentiation serves to give added weight to the speaker’s endorsement of p , which is qualified from the outset as subjective and potentially non-consensual. It is unsurprising for such uses to be frequent in political debate, as Simon-Vandenberg (2000) has noted, since it is in this genre of discourse that speakers stake out their personal position most explicitly. This corresponds to Kaltenböck’s (2010) “booster” function.

4.5 Configurations with final ‘I think’

Earlier again we noted the salience of numerical expressions in utterances featuring final ‘I think’. In fact it is not numerical expressions as such but rather expressions which are presented as potentially objectifiable. The fact that ‘I think’ is in final position means that this is a two-step process. An initial representation which might just as well be attached to the speaker as to any other enunciative source, presented as a “statement of fact”, is retrospectively adjusted to mark the potential for differentiation between sources, and a concomitant “weakening” or “distancing” effect.

17. The asterisk is used conventionally within the TEPO to indicate an indeterminate or “wild-card” value. Elsewhere I argue for a distinction between deliberative ‘I think’ in evaluative contexts, opposing the speaker and a co-speaker, and tentative ‘I think’ in assertive contexts, opposing the speaker and a generic enunciative instance, the speech community (Ranger 2018).

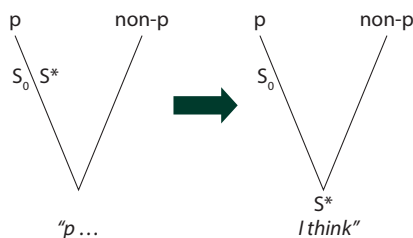


Figure 2. A two-step process dissociating the speaker's endorsement S_0 of p from that of other potential enunciative sources S^*

Note how this operates in Examples (15)–(19), which provide typical instances featuring proper nouns, dates, quantities, etc.

- (15) I dunno, it was either *Clive or Richard* I think. KE6 5293
- (16) Nineteen squared's *three hundred and ninety ninety one* I think. FMM 618
- (17) ... there were three or two major epidemics in Oxford, of what they call plague, but it was probably a form of typhus, *in 1643 and 44*, and a good deal of sickness, I think, still *in 1645* ... KRN 152
- (18) Well he's still *in Manchester* I think. KC9 1727
- (19) *Two pounds, thirty* I think. KBE 10011

In these cases the responsibilities for an initial representation are reduced from potentially any speaker – since here the speaker first appears merely to be reporting “facts”¹⁸ – to the speaker alone. In keeping with our corpus data, Kaltenböck (2010: 249), who calls the function of ‘I think’ in examples such as these an “approximator” or “shield”,¹⁹ notes the preference for numerals and similar terms of precision here:

[...] numerals, days of the week/month, etc. lend themselves particularly well to approximation. [...] On a more general level, there is also a close link between approximation and precision of categorial delimitation in the sense that approximative use presupposes clear category boundaries. This is why referring terms and predicates are the natural environment for approximators or bushes [...].

18. The concept of the “fact” is not a linguistic reality, of course, but there are certain types of propositional content which can be seen as independent of their endorsement by a subject and others which on the contrary exist only by virtue of this endorsement.

19. As mentioned previously, the difference between the two appears to be essentially pragmatic and interpretative: approximators play on referential values, shields on speaker / co-speaker relationships. If too bold an assertion represents a potential threat to a co-speaker, then a “approximator” might be seen as a “shield”.

4.6 Configurations with medial 'I think'

The values of 'I think' when initial or final are largely dictated by its position, as seen above. When 'I think' is medial, however, three possibilities seem to be available.

- 'I think' is used with a potentially oppositional effect, and clausal scope, as a way of reinforcing speaker endorsement, as when it is initial.
- 'I think' is used with a potentially distancing effect, and non-clausal scope, as a way of limiting the endorsement of a particular term (often a referring item) to the speaker alone, as when it is final.
- 'I think' is also claimed to function as a "filler", enabling the speaker to hold the floor in cases of disfluency.

Let us look at each of these in turn.

The first of these, in which 'I think' reinforces speaker endorsement, might be illustrated by (20)–(23) below.

- (20) Rupert *does*, *I think* [pause] tend to get through a hell of a lot of milk. KBL 3213
- (21) Ergo, we *have to look*, *I think*, at the record of this company over the last few years HM6 78
- (22) And *it's nice*, *I think*, to include [pause] this [pause] plough, if that's what it is. HM2 453
- (23) As I understand, *it is not I think* what I voted for in number three JS9 765

In these examples, 'I think' gives added weight to the assertion, and also contributes to the thematic organization of the utterance, with topic to the left and comment to the right, topicalizing the speaker's commitment to a particular option, in the face of potential opposition.²⁰ In (20), for example, the elements to the left of 'I think', 'Rupert does', indicate the speaker's commitment to the following predication. In (21), medial 'I think' places the speaker's expression of obligation 'we have to look' in topic position. In (22) and (23) the topicalized terms are 'it's nice' and 'it is not' respectively.²¹

In the next set of examples, medially placed 'I think' takes hedging values.

20. Ziv (2002) considers the linking role of parentheticals in medial position.

21. Simon-Vanderbergen (2000: 49–50) points out the use of medial 'I think' in thematic organisation in political discourse.

- (24) Er, I did go on H R T because of the osteoporosis risk, but I was unlucky I was one of the [pause] ‘I think’ twelve percent who had developed breast trouble and had to come off it. FL4 152
- (25) [...] she had some friends in Germany and of course they, she said they were hoarding up just at the borders of Poland ‘I think’ it would be then. GYW 405
- (26) Sir er [pause] you hinted that er if there are to be further written representations you’d be looking to directing a cutoff date, and you hinted at ‘I think’ tomorrow as being one possibility FMP 943

In (24)–(26), ‘I think’ is clause-medial. This classification is somewhat misleading, however, since its scope is in fact non-clausal, as it targets respectively the numerical precision ‘twelve percent’, the place name ‘Poland’ and the date ‘tomorrow’. In the first two cases, we might see ‘I think’ as an approximator, as the speaker recognises a possibility of imprecision. In the third case this acknowledgement has clearer pragmatic consequences (as a “shield”, “softener”, etc.), since the original source of the reference ‘tomorrow’ is apparently the co-speaker ‘you’. In using ‘I think’, together with ‘hinted’ and ‘one possibility’, the speaker reminds the co-speaker (note the honorific ‘sir’) of his or her commitment to a ‘cutoff date’ while at the same time allowing a certain diplomatic margin for error adjustment.

Kaltenböck’s (2009a) study of prosodic features of such occurrences shows that clausal scope is the unmarked option and that when ‘I think’ has non-clausal, or phrasal scope, this is invariably marked in spoken corpora by prosodic binding between the comment clause and its scope (Kaltenböck 2009a: 69–70, for example).

The purported use of medial ‘I think’ with a structural or filler function (Kaltenböck 2009a: 71) is very hard to separate clearly from other values. Kaltenböck gives the following example, pointing to the association of ‘I think’ with other fillers (‘I mean’, ‘really’), pauses and repetitions as proof of its use in contexts of disfluency:

- (27) I mean *I think* really uhm <,,> it’s very difficult to to to produce any form of art unless you are driven (Kaltenböck 2009a: 71)²²

There is however nothing to prevent one seeing (27) as a case of ‘I think’ being used to reinforce speaker endorsement in potential opposition to other enunciative sources, associated with markers of commitment such as ‘I mean’, ‘really’ or ‘very’. Similarly, if ‘I think’ in (28) exists in a context of disfluency, it also acts as a hedge, in allowing for potentially discordant views on a point of numerical precision.

22. The double comma <,,> indicates a long pause, in the notation used in Kaltenböck (2009).

- (28) Erm *I think* we was there about nine months. It wasn't a bad place but er a very small one you know. Erm from then we got a council house which is where the wife's living at the moment. Erm <pause> *I think* <pause> *I think* we'd have been there about No. About a year when we decided to split up. HEN 22–26

In short, it seems unnecessary to consider that 'I think' possesses a structural or filler function separate from its other values, since these other values are also demonstrably present, even when 'I think' occurs in contexts of disfluency.

4.7 Summary

To sum up, the schematic form of 'I think' indicates that a target notion p is located relative to a subjective representation of the speaker S_0 (in the notation of the TEPO $\langle p \in S_0 \rangle$). Contextual elements, including position, further determine this subjective localisation relative to other speakers, S^* . When 'I think' is initial, and in association with evaluative items, the speaker's endorsement of p is fully assumed and in opposition to other alternative views ($\langle S_0 \neq S^* \rangle$). When 'I think' is final, frequently in association with referring terms, the speaker indicates subjective endorsement retroactively, with a concomitant distancing effect, as what first appeared as objectifiable fact, is assigned to the speaker alone (movement from $\langle S_0 = S^* \rangle$ to $\langle S_0 \omega S^* \rangle$). When 'I think' is medial, both configurations are possible: lexico-grammatical and prosodic parameters enable disambiguation.

5. Discussion and conclusion

From the above discussion it should be clear that in the enunciative approach of the TEPO, 'I think' is neither an evidential, nor an epistemic discourse marker. Evidentiality can be understood either as a grammatical category or as a linguistic category (Culioli et al 1981). As a grammatical category, evidentiality involves the indication of the source of information. As a linguistic category, evidentiality involves the obligatory expression of the corresponding grammatical category *via* specific markers known as evidentials. The marker 'I think' does involve indication of the information source – albeit by default, as the “information” – or rather the representation corresponding to p in our model – is located relative to the speaker. There is nothing obligatory about this, however: the evidentiality expressed in 'I think' is at most a means of rendering explicit what, in languages with evidential systems, would be expressed by specific morphemes, etc. Similarly, 'I think' does not express epistemic uncertainty. In some cases – typically when placed finally – it

may allow us to infer uncertainty, but in others, it may on the contrary add extra commitment to the expression of a point of view.

Various contextually situated values of 'I think' can be shown to result from the interaction of (1) a schematic form, indicating that a notion *p* is located relative to a subjective representation, (2) the scope of the locator 'I think', which is closely related to considerations of position and prosody, and (3) the lexico-grammatical properties of the target. Variations in these parameters contribute to subjective positioning of the speaker relative to other potential enunciative sources according to two fundamental templates. Either (1) the speaker's endorsement of *p* frames the utterance, against potentially discordant enunciative positions, in an operation of differentiation, corresponding to boosting, reinforcement or deliberative values, in other models, or (2) the speaker's endorsement of *p* operates retroactively, mitigating inferences of objective authority and opening the way for potential opposition, corresponding to hedging, shielding or tentative values, in other models. It would be more appropriate to consider these, not as values of the sequence 'I think', so much as values derived from specific macroconstructions involving this string together with constraints of position, context, prosody, etc. Provided meanings are constructed in terms of complex configurations of abstract forms, it becomes unnecessary to postulate new and original functions for 'I think'.

I hope also to have shown how the identification of those parameters salient in the configuration of abstract forms may be facilitated and given quantitative support with the use of targeted corpus queries. In this perspective the collocations of a marker do not merely contribute descriptive statistics about the usage of the marker but additionally provide precious indications for the enunciative linguist to model the abstract schematic form associated with the marker itself.

References

- Aijmer, Karin. 1997. 'I think' – An English modal particle. In *Modality in Germanic Languages*, Toril Swan & Olaf J. Westvik (eds), 1–48. Berlin: Mouton de Gruyter. doi:10.1515/9783110889932.1
- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: OUP.
- Brinton, Laurel J. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions* [Topics in English Linguistics 19]. Berlin: Mouton de Gruyter. doi:10.1515/9783110907582
- Brinton, Laurel J. 2008. *The Comment Clause in English: Syntactic Origins and Pragmatic Development*. Cambridge: CUP. doi:10.1017/CBO9780511551789
- Chafe, Wallace L. & Nichols, Johanna. 1986. *Evidentiality: The Linguistic Coding of Epistemology*. Norwood NJ: Ablex.
- Culioli, Antoine & Desclés, Jean-Pierre. 1981. *Systèmes de représentations linguistiques et méta-linguistiques. Les catégories grammaticales et le problème de la description de langues peu étudiées*. ERA 642.
- Culioli, Antoine. 1990. *Pour une linguistique de l'énonciation*, Vol. 1. Gap: Ophrys.

- Culioli, Antoine. 1999. *Pour une linguistique de l'énonciation: Formalisation et opérations de repérage*, Vol. 3. Gap: Ophrys.
- Fletcher, William H. 2003. *Phrases in English*. <<http://phrasesinenglish.org/>>
- Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hoffmann, Sebastian, Evert, Stefan, Smith, Nicholas & Lee, David. 2008. *Corpus Linguistics with BNCweb: A Practical Guide*. Frankfurt: Peter Lang.
- Holmes, Janet. 1990. Hedges and boosters in women's and men's speech. *Language & Communication* 10(3): 185–205. doi:10.1016/0271-5309(90)90002-5
- Kaltenböck, Gunther. 2009a. English comment clauses: Position, prosody, and scope. *Arbeiten aus Anglistik und Amerikanistik* 34(1): 49–75.
- Kaltenböck, Gunther. 2009b. Initial 'I think': main or comment clause? *Discourse and Interaction* 2(1): 49–70.
- Kaltenböck, Gunther. 2010. Pragmatic functions of parenthetical 'I think'. In *New Approaches to Hedging*, Wiltrud Mihatsch, Stefan Schneider & Gunther Kaltenböck (eds), 237–266. Leiden: Brill. doi:10.1163/9789004253247_012
- Kärkkäinen, Elise. 2003. *Epistemic Stance in English Conversation: A Description of its Interactional Functions, with a Focus on 'I think'* [Pragmatics & Beyond New Series 115]. Amsterdam: John Benjamins. doi:10.1075/pbns.115
- Lakoff, Robin. 1975. *Language and Woman's Place*. New York NY: Harper and Row.
- Lohmann, Arne & Koops, Christian. 2016. Aspects of discourse marker sequencing: Empirical challenges and theoretical implications. In *Outside the Clause. Form and Function of Extracausal Constituents* [Studies in Language Companion Series 178], Gunther Kaltenböck, Evelien Keizer & Arne Lohmann (eds), 417–446. Amsterdam: John Benjamins.
- Palander-Collin, M. 1999. *Grammaticalization and Social Embedding: 'I think' and Methinks in Middle and Early Modern English*. Helsinki: Société Néophilologique.
- Ranger, Graham. 2018. *Discourse Markers: An Enunciative Approach*. London: Palgrave Macmillan.
- Ross, John R. 1973. Slifting. In *The Formal Analysis of Natural Languages*, Maurice Gross, Morris Halle & Marcel-Paul Schützenberger (eds), 133–172. The Hague: Mouton.
- Simon-Vandenberg, Anne-Marie. 2000. The functions of 'I think' in political discourse. *International Journal of Applied Linguistics* 10(1): 41–63. doi:10.1111/j.1473-4192.2000.tb00139.x
- Thompson, Sandra A. & Mulac, Anthony. 1991. A quantitative perspective on the grammaticization of epistemic parentheticals in English. In *Approaches to Grammaticalization*, Vol. 2: *Types of Grammatical Marker* [Typological Studies in Language Vol. 19(2)], Elizabeth Closs Traugott & Bernd Heine (eds), 313–329. Amsterdam: John Benjamins. doi:10.1075/tsl.19.2.16tho
- Urmson, James Opie. 1952. Parenthetical verbs. *Mind* 61(244): 480–496. doi:10.1093/mind/LXI.244.480
- Van Bogaert, Julie. 2010. A constructional taxonomy of 'I think' and related expressions: accounting for the variability of complement-taking mental predicates. *English Language and Linguistics* 14(3): 399–427. doi:10.1017/S1360674310000134
- Van Bogaert, Julie. 2011. 'I think' and other complement-taking mental predicates: a case of and for constructional grammaticization. *Linguistics* 49(2): 295–332.
- Ziv, Yael. 2002. This, I believe, is a processing instruction: Discourse linking via parentheticals. In *Proceedings of Israel Association for Theoretical Linguistics 18*, Yehuda N. Falk (ed.). <<http://linguistics.huji.ac.il/IATL/18/Ziv.pdf>>

Embedding evidence in Tagalog and German

On two types of evidentials

Jennifer Tan and Johannes Mursell

Goethe University Frankfurt / ILLA-CSIC

We compare two particles with similar meaning contributions from two different languages, German *wohl* and Tagalog *yata*, both roughly translatable as ‘I infer’. Based on their context usage, we claim that both are inferential evidentials. Despite their identical meaning, we assume a non-uniform treatment of evidentials, in view of their differing behaviors. By applying well-established tests, we determine the level of meaning they operate on, the propositional level for *yata* and the illocutionary level for *wohl*. Our data show that the embeddability test needs modification, allowing for illocutionary operators to be embedded, but only in root-like clauses, as *wohl* does. *Yata*, being embeddable in both root and non-root clauses, operates on a propositional level, and is thus a modal evidential.

Keywords: evidentiality, epistemic modality, inferential, (non-)root clauses, Tagalog

1. Introduction

Ever since the introduction of the term ‘evidentiality’ in 1947 by Franz Boas (apud Jacobsen 1986), the linguistic expression of source of information across languages has been investigated at length. Scholars have been concerned with many different aspects of this linguistic notion, such as its domain, its typology, grammaticalization of evidentials, their relation to other categories, etc. (for an ample overview of the many pertinent issues to evidentiality, see Aikhenvald 2004; Aikhenvald 2018). Here we will take into consideration the issues of what evidentials are (and what not) and how they operate in a sentence.

Regarding what they are, the function of evidentials is to convey the speaker’s evidence for their claim. Definitions invariably refer to them as the linguistic encoding of source of evidence but vary in other aspects such as whether or not they

additionally encode degree of certainty or commitment (Brugman & Macaulay 2015) towards the content of the utterance. Speaker's degree of certainty towards the truth of the prejacent, that is, the propositional content, is usually encoded by epistemic modals. The particles at hand, the Tagalog *yata* and the German *wohl*, have been described to express uncertainty towards the content of their host utterance (Schachter & Otnes 1972; Zimmermann 2004, respectively), thus they appear to express only part of the typical evidential meaning, namely degree of certainty/commitment. Therefore, at first sight they may be argued to convey epistemic uncertainty alone. However, on the grounds of context usage, we argue here that such uncertainty stems from the limited evidence available to the speaker. Hence, we propose that these particles necessarily convey the information source as well, since their felicitous use requires some source of information on which the speaker may infer the content of their utterance. Thus, accordingly, they should clearly be considered evidentials, more specifically, inferentials.

As for how evidentials operate in a sentence, recent research among semanticists has dealt with their scopal behavior, semantic content, (non)-at-issueness, or the level on which they contribute their meaning (see McCready 2015a). In general, evidentials may be classified based on the level they operate on, either on a propositional level (e.g. Izvorski 1997; Matthewson et al. 2007; Faller 2011; McCready & Ogata 2007; Peterson 2010; Waldie, Peterson, Rullmann & Mackie 2009) or on an illocutionary level (Faller 2002). The main assumption behind this distinction is that the former operator type contributes to at-issue content, whereas the latter does not. Among the tests that have been proposed to distinguish them are those concerned with truth values and those involving scope and embeddability (see overview in Waldie et al. 2009, cf. Korotkova 2016 for the often heterogeneous behavior of evidentials and, hence the problematic assumption of this dichotomy). Here, we will take a closer look at the embeddability test, that is, whether the evidential can be interpreted as part of the propositional content of an embedded clause or not, by contrasting the environments in which *yata* and *wohl* may occur. This test has been argued to show differences between evidentials operating on a propositional level, which can be embedded rather freely, and those operating on an illocutionary level, which have been claimed to not embed at all (Faller 2002). We propose a refinement of the embeddability test that enables us to distinguish the two types of evidentials, inasmuch as illocutionary operators or speech act operators (SAOs henceforth) are dependent on the presence of illocutionary force in the clause they occur in and therefore are disallowed in non-root clauses (following Haegeman 2006; Coniglio 2008), but can be embedded in root-like clauses. Thus, contrasting the behavior of *yata* and *wohl* in non-root clauses allows an in-depth investigation of their respective levels of operation.

This chapter is structured as follows. Section 2 deals with the concept of evidentiality and its relation to epistemic modality. We introduce the elements in question, Tagalog *yata* and German *wohl*, and provide examples tested in different contexts supporting the claim that both are inferential evidentials. In Section 3, we outline the differences between evidentials operating on a propositional level or modal evidentials (Matthewson 2012, 2015) on the one hand, and those that operate on an illocutionary level (SAOs) on the other hand. We check the standard contrasting diagnostics' applicability to both particles and pay special attention to their embeddability in (non-)root clauses. These indicate that *wohl* is a SAO, for it seems to be allowed only in root clauses, whereas *yata* is a modal evidential, having less restrictions with respect to embeddability. Section 4 concludes.

2. Evidentiality: *Yata* and *wohl* as evidential markers

In this section, we discuss the notions of evidentiality and epistemic modality, in support of a view of evidentials as indicating both source and reliability of information. Then, we justify our claim that *wohl* and *yata* are inferential evidentials, based on their meaning contribution and context usage.

2.1 The notion of evidentiality

Evidentiality has been mostly defined as the linguistic encoding of source of evidence. Evidentials may come in the form of affixes, verbs, modals, clitics or particles (Aikhenvald 2004). While most studies on evidentials agree on defining them as conveying the information source for the propositional content of an utterance, there is discrepancy on whether or not it may encode degree of certainty as well (e.g. Dendale & Tasmowski 2001; Brugman & Macaulay 2015). The expression of degree of certainty is traditionally relegated to the domain of epistemic modals. There is an obvious connection between the two domains, given that a speaker's source of information for a given claim is correlated with the certainty with which they believe such claim. While some scholars argue that only elements that strictly encode the speaker's source of information should be considered evidentials, thus fully distinguished from other related categories such as modality or mood (de Haan 1999, 2001; Aikhenvald 2004; Nuyts 2006), recent research corroborates the need for an inclusive definition of evidentiality (as described by Chafe & Nichols 1986; Givón 2001, a.o.), whereby evidentials indicate both source and reliability of information (Rooryck 2001: 125). We assume this definition of evidentiality, considering the delimitation between the two categories is often imperceptible as

strongly suggested by numerous crosslinguistic studies on evidentials. According to Brugman & Macaulay (2015), evidentials consistently express the source of evidence, belong to a closed class of grammatical items, and may vary in obligatoriness, complementarity of meaning of items, truth-conditionality, shift of origo to 3rd person, and particularly, degree of certainty, commitment or informativity. Likewise, Matthewson (2015) proposes that evidentials may encode one or more dimension, with a direct or indirect value in each dimension. The three dimensions are: (a) evidence type, either firsthand (sensory information) or secondhand (reports or reasoning); (b) evidence location, involving the speaker witnessing directly the event expressed in the preajacent or indirectly; (c) evidence strength, concerned with trustworthiness or reliability of the evidence. Thereby, recent literature has referred to evidentials potentially expressing reliability (see McCready 2015b). Formally, the inextricable relation between both categories is straightforwardly reflected in analyses of evidentials as modals and vice-versa, which we explore in Section 3.

Moreover, the inclusive definition of evidentiality assumed in this study receives further empirical support when looking into a specific kind of evidentials, namely inferential evidentials. The general consensus since Palmer (1986) (see also Dendale 1994, 2001; Cornillie 2009; Barbet 2012) seems to be that the notions of inferential evidentiality and epistemic modality are indistinct, considering that in using either epistemic modals or inferentials, there are always reasons for the speaker to assert the (im-)possibility of a given proposition. After all, modal judgments are based on some type of evidence (Rooryck 2001). However, not all epistemic modals can be regarded as inferential evidentials, as the latter necessarily involve an inference based on observable results or mental reasoning (Willett 1988). Accordingly, von Stechow & Gillies (2010) point out the evidential nature of a strong necessity modal such as ‘must’, signalling indirect inference, in contrast to other epistemics such as ‘probably’ or ‘might’, which do not have such a requirement (cf. de Saussure 2014 for whom epistemic and evidential meanings of modal markers arise pragmatically).

Based on this theoretical background and discussion, we investigate the question of whether or not the German *wohl* and the Tagalog *yata* may be characterized as evidentials and if so, their evidence type. *Wohl* and *yata* are said to express epistemic uncertainty, which should make them epistemic modals. However, as we will show in detail, not only do they express a degree of certainty, but also an inferential type of evidence, considering that their use requires for the speaker to base their inference on a certain piece of evidence, which supports an inferential evidential treatment of both particles, as we will see in the following section.

2.2 Wohl and *yata* as evidentials

While some of the data provided here may be taken from the references cited, by default, the data presented in this paper were constructed by the authors, based on our intuitions as native speakers, and then presented to other native speakers of each language for direct implication judgments and acceptability tasks. The speakers were asked questions related to the context and to judge whether the utterances are felicitous in the given context (Tonhauser & Matthewson 2015). The Tagalog-speaking consultants were born and raised in Manila, two men and two women past their 50s, with a university level of education and no prior linguistic training. The German-speaking consultants included 5 people raised in several regions of Germany, varying in age from 30 to 59, with all having received a university education. Part of the data was taken and/or adapted from two different corpora: (a) corpus of speeches from the German parliament, built and annotated as part of the SFB 632 project A8, funded by the DFG; (b) SEALang Library Tagalog Text Corpus, which includes more than two million words taken from the Ramos *Tagalog-English Dictionary* and the 200,000 words *Tagalog Literary Text* collection prepared by the *Philippine Languages Online Corpora* project.

2.2.1 *Yata*

The Tagalog *yata*, unlike *wohl*, has so far not been discussed in the literature from a theoretical point of view (Tan 2016). Schachter and Otanes (1972: 428) in their reference grammar of Tagalog merely state [*yata* is an enclitic particle] used in statements (not in questions or imperatives) to express uncertainty or lack of conviction”. It is usually translated as an adverbial *maybe* or *perhaps*, or a propositional attitude predicate such as *think* or *seem*.

- (1) *Umulan yata kahapon.*
 rained YATA yesterday
 ‘I think it rained yesterday’.

However, the translation of evidentials is always complex and often misleading. As a particle, a verbal translation like the one in (1) can be mistaken for some sort of embedding element. Hence, it seems syntactically inaccurate, even if main clauses such as *I think* in (1), where the embedded clause has main point status, have been argued to have the pragmatic function of an evidential (see Simons 2006). Moreover, the adverbials *maybe* or *perhaps*, conveying epistemic possibility, cannot reflect the *bona fide* meaning contribution of *yata*, as we can see from the contrast in felicity of the utterances in (2) for such context, marked by #.

- (2) Context: I did not go out yesterday, so I cannot know for sure:
 {*Maybe/Perhaps*} *it rained yesterday*.
 #*Umulan yata kahapon*.
 rained YATA yesterday¹
 ‘It rained yesterday (I infer)’.

The contribution of *yata* differs from the adverbials *maybe* and *perhaps* in that it requires an indication of indirect inference, similarly to *must* (von Stechow & Gillies 2010). Consultants reject the given context, since the speaker has no reason to infer *p*: the context targets solely the speaker’s uncertainty, but rules out any type of evidence on which they may base *p*. A context of mere uncertainty calls for the use of an epistemic modal *baka* ‘maybe’. Hence, *yata* needs a piece of evidence based on which the speaker can infer the propositional content ‘it rained yesterday’, either through reasoning or some observable result (e.g. the speaker sees that the ground is wet).

Regarding *yata*’s category, it belongs to a group of eighteen second position (2P) particles that express an array of meanings including temporal, social, interrogative, modal and evidential. The syntactic complexity of these 2P particles has been the subject of numerous studies (e.g. Schachter 1973; Kroeger 1998; Anderson 2008; Kaufman 2010), but we do not deal with this issue here. We only want to point out how *yata*, as a 2P particles, occurs as the second element in the clause, right after the predicate, for Tagalog is a predicate-initial language (Kroeger 1993).

As pointed out by Schachter & Otnes (1972), *yata* is restricted to statements. It is thus disallowed in interrogatives and imperatives. It appears in complementary distribution with the speculative particle *kaya*, which occurs only in such clause types and not in statements (Tan 2016). *Yata* does not show origo shift, that is, the holder of the evidential information does not change (in terms of Garrett 2001; see Brugman & Macaulay (2015: 216–222) for a crosslinguistic survey of evidential shift instances) except when embedded. Thereby it usually shows anchoring to the speaker, but in the embedded occurrences, it deictically refers to the subject of the matrix clause, whose inference is being reported by the speaker of the subsequent utterance that refers to it. Otherwise, *yata* is not able to undergo shift, and since it cannot occur in questions, it does not participate in interrogative flip (in terms of Speas & Tenny 2003). Intuitively, questioning one’s own inference seems to yield an uncertainty that may belong to the realm of speculation, that is, the speaker wonders about the possibility of the described event in the utterance. Speculation is lexically expressed by *kaya*, which is anchored either to the speaker

1. Used abbreviations: EXIS existential, GEN genitive, LNK linker, LOC locative, NOM nominative, PERS personal, PL plural, SG singular.

or the addressee, as can be seen in the readings in (3a). The utterance with speaker origo functions as a rhetorical question, as one wonders to themselves. When anchored to the addressee, the speaker expects the hearer to not have access to direct evidence for the described event and therefore will be inferring their answer. As for imperatives, the use of *kaya* expresses that “the speaker speculate[s] about the desirability of the action” (Schachter & Otanes 1972: 426). The example in (3b) shows the impossibility of *yata* in the imperative, for a speaker cannot make a claim of an inference in a command. Thus, it is only logical that declaratives, serving to convey a statement, host *yata*.

- (3) a. *Umulan* {**yata/ kaya*} *kahapon?*
 rained YATA KAYA yesterday
 Speaker origo: ‘Did it rain yesterday, I wonder?’
 Addressee origo: ‘Did it rain yesterday, what do you think?’
- b. *Kumain na* {**yata / kaya*} *kayo.*²
 eat already YATA KAYA 2PL
 ‘Perhaps you should eat’.

Provisionally, for its detailed analysis goes beyond the scope of this chapter, we simply advance that *yata* semantically and pragmatically shows a modal behavior, as we will see in Section 3.1.

2.2.2 *Wohl*

The German *wohl* has traditionally been analysed as a discourse particle (DiP henceforth).³ DiPs, sometimes also called modal particles (for example in Abraham & Leiss 2009), are a small class of words that generally express the speaker’s and/or hearer’s attitude towards the content of the utterance. More precisely, following Thurmair (1989) and much subsequent work, we assume that DiPs relate the content of the utterance they occur in to the Common Ground, the shared knowledge of the interlocutors.⁴ DiPs have been identified in several languages (e.g. Zeevat

2. The imperative in Tagalog may be realized as the base form of the verb, e.g. *kain* ‘to eat’, or the agentive voice, here *kumain*. It is certainly possible to utter *Kumain na yata kayo*, but it is then interpreted as an assertion ‘You have already eaten (I infer)’, whereas **Kain na yata kayo*, unambiguously imperative, is not allowed.

3. We opt for the term discourse particle, DiP, here for several reasons. Even though we believe that this term is preferable in general since it avoids the arguably wrong allusion to modality in its linguistic meaning, we especially employ it in this paper since it nicely contrasts with the category of modal evidentials. However, we are well aware of the controversial debate surrounding this issue.

4. The term Common Ground goes back to a lecture of Paul Grice from 1967, see Stalnaker (2002) for an exposition.

2000 for Dutch; Scherf 2017 for Swedish; Coniglio & Zegrean 2012 for Italian and Romanian; and for a more traditional analysis of German DiPs in the spirit of Helbig 1988 and Hentschel 1986, see Diewald 2006). However, due to their complex properties, they have not received a uniform treatment in the literature. In general, speech act restrictions exist for all DiPs, such that no speech act can host all possible DiPs.

In this chapter, we focus on one specific DiP in German, *wohl*, which, roughly speaking, expresses that the speaker is unsure about (parts of) the content of the utterance. Syntactically, DiPs in German clauses are restricted to the middle field, that is, the position between the finite verb in second position in main clauses, or the complementizer position in subordinate clauses, and the sentence final position usually marked by verbal particles in main clauses or the finite verb in embedded clauses (*modulo* extraposition). Exceptional occurrences of DiPs as part of nominal phrases, or complex *wh*-constituents are possible, but not relevant for our purpose. As mentioned, *wohl* expresses uncertainty towards (parts of) the content of the utterance. In declaratives, this is speaker uncertainty, while in questions, the uncertainty expressed by *wohl* can undergo interrogative flip, so the uncertainty can be shifted to the addressee, in contrast to *yata*

- (4) a. *Peter hat wohl ein Bier getrunken.*
 Peter has WOHL a beer drunk
 ‘Peter drank something, I infer it was beer’.
- b. *Was ist wohl die Hauptstadt von Tansania?*
 What is WOHL the capital of Tanzania
 ‘What is the capital of Tanzania, what do you infer?’

Concerning the contribution of the DiP in detail, several approaches have been presented in the literature. Thurmair (1989: 139ff) analyses the DiP as contributing the feature <EINSCHRÄNKUNG> (‘restriction’), restricting the validity of the utterance. This serves as an explanation for the datum in (5), in which *wohl* seems to lead to infelicity. Since the utterance presents a universally known fact, expressing uncertainty towards the expressed proposition is infelicitous as in (5).⁵

5. Note that *wohl* in (5) does have a felicitous use, however not as DiP. Nearly all German DiPs have homophonous counterparts that serve different functions in the language. A different use of *wohl*, which goes back to the same word as English *well*, is to show a contrast with a preceding statement. In this case, which usually goes along with stress on *wohl*, (5) is indeed felicitous but it is not possible to treat *wohl* as DiP. Having unstressed *wohl* in (5), and thus an unambiguous DiP use, leads to infelicity. In this paper, we are only concerned with the DiP use of *wohl*. Thanks to a reviewer for reminding us to mention this contrast.

- (5) #*Die Erde ist wohl rund.*
 the earth is WOHL round
 Intended: ‘The earth is round, I infer’.

Zimmermann (2004) presents a different analysis of *wohl*, assuming that it expresses a certain amount of epistemic uncertainty towards the expressed descriptive meaning. Based on the theory of common ground by Stalnaker (2002), in which the utterance of *p* is intended to introduce *p* into the common ground, *wohl p* only introduces the assumption *that p* or *assume(p)* into the common ground. Since *wohl* expresses epistemic uncertainty, according to Zimmermann (2004) it is necessarily restricted to speech acts which operate on epistemic modal bases, which are declaratives and questions.

- (6) $\llbracket \text{wohl} \rrbracket = \text{ASSUME}_x(p)$ (with $x = \text{speaker, hearer, or both}$)

Gutzmann (2015) presents yet another possible analysis based on the notion of expressive meaning (Potts 2007) in which DiPs operate on an independent level of meaning by contributing *use conditions* to the utterance in contrast to *truth conditions*. As we will show when discussing our analysis of *wohl* as speech act operator (SAO) in 3.2, taking DiPs to operate on an independent semantic level cannot sufficiently account for *wohl*'s behavior.

Even though the precise analysis of the contribution of *wohl* to the clause is of secondary importance for our point, we follow Egg & Mursell (2017) in assuming that *wohl* can target and negate specific felicity conditions of speech acts (Searle 1969; Vanderveken 1990), with the speech act being linguistically encoded as part of the sentence in one way or the other. The specific felicity conditions negated by *wohl* depend on the speech act. In assertions, *wohl* negates the first preparatory condition which states that usually the speaker has enough evidence to assert *p*. This lack of evidence then derives the intuition that *wohl* expresses uncertainty towards the content of *p*. In questions, *wohl* targets the third preparatory condition, namely that the speaker believes the hearer to have the answer to the posed question, and negates it. Thus, the uncertainty is shifted to the hearer who receives permission to answer the question without having sufficient evidence for the response.

Treatments of *wohl* as a proper evidential can also be found in the literature. Based on evidence from translation, Haumann and Letnes (2012) argue that *wohl* does indeed display evidential properties. They claim that at least certain variants of *wohl* behave like a weak relative of proper evidential adverbs like *offensichtlich* ‘apparently’ (*weak* in the sense of Cardinaletti & Starke 1999). Since translations of DiPs into other languages can only offer a very restricted picture of their function, our chapter supports their assumption by using felicity judgements from native speakers as a basis to maintain the claim of *wohl* being an evidential. Similarly,

Modicom (2012) also argues that *wohl* contains an inherent albeit covert inferential component. He claims that *wohl* expresses a certain amount of speaker commitment towards the content of the utterance. In combination with a lack of direct evidence for the expressed content, it gives rise to an inferential meaning component in addition to a weak commitment reading. Modicom further argues that neither of the two meaning components, weak commitment or inferentiality, are part of the core meaning, and both arise either as implicature or in combination with an unsubstantiated proposition. In her analysis, Repp (2006, 2009) does not directly focus on *wohl* as an evidential, but argues for the modification of degrees of strength to be at the core of *wohl*'s contribution to the clause. Following Vanderveken (1990), Repp assumes that a neutral assertion carries a neutral degree of strength as part of the sincerity condition of its assertive declarative force. Certain operators can modify this degree of strength, and *wohl* in particular lowers it. Since degree of certainty or strength of commitment are core properties of evidentials across languages, Repp's approach can be taken to at least partially support an evidential analysis of the DiP *wohl*. Recently, Eckardt (2017) related to the inferential component of *wohl* by proposing that it marks non-monotonic inference, that is, use of *wohl* shows that, according to what the speaker knows, *p* holds true, although further evidence may prove otherwise.

Wohl's contribution of uncertainty as stated in the literature has already been outlined above, but interestingly, this is not its only meaning component. German native speakers reject *wohl* in (7), which is parallel to the Tagalog Example (2) above.

- (7) Context: I did not go out yesterday, so I cannot know for sure:
 #*Es hat gestern wohl geregnet.*
 it has yesterday WOHL rained
 'It rained yesterday, I infer.'

The context in (7) is the decisive factor. Similar to *yata*, if the speaker does not have any kind of evidence for their assumption, *wohl* cannot be used felicitously. However, changing the context slightly by providing the speaker evidence for his assumption, that is, the weather report, water in the garden, or such, makes the utterance in (7) felicitous. From this, we conclude that *wohl*, like *yata*, also requires some indirect kind of evidence. In the next subsection, we will further substantiate the claim that both *yata* and *wohl* should be treated as evidentials by introducing more contexts in which they appear to behave similarly.

2.2.3 Context usage

Despite occurring in completely unrelated languages, *wohl* and *yata* do not only share the requirement for an inferential type of evidence, they also contribute very similar meanings in those contexts in which both can be used. The contexts in the following examples are set to target epistemic necessity (adapted from van der Kloek 2014). Just like the English *must* (von Fintel & Gillies 2010), *wohl* and *yata* seem to convey strong epistemic necessity. The speaker's inference is made through mental reasoning (as in (8)) or via observable results (as in (9)).

- (8) The class goes from 3 to 4 p.m. and it is already 3.20 p.m. The professor is not here yet and is never late.
- a. *Sie kommt heute wohl nicht mehr.*
she comes today WOHL not more
'She is not coming anymore today (I infer)'.
 - b. *Hindi na yata siya darating nyagon.*
not anymore YATA (s)he will.come today
'She is not coming anymore today (I infer)'.

The context in (8) targets necessity: considering the class lasts an hour, the professor is punctual and she is already late, it is necessarily the case that the professor is not coming at all. Consultants reject the use of the particles in a minimally different context targeting epistemic possibility, in which the professor is known to be inconsistent and tardy. Habits such as tardiness seem to not constitute enough evidence to infer *p*. The same holds for (9). Since the speaker has already looked for their keys everywhere and given that some noise of keys can be heard inside of the bag, it is necessarily the case that they are there. Consultants reject the utterances in (9a) and (9b) when given a slightly different context in which the speaker has not looked for the keys everywhere, leaving the possibility of the keys being somewhere else, or if the speaker has a habit of leaving them in the bag. These data show that both particles share expression of epistemic necessity and do not discriminate inference based on either reasoning or observable results.

- (9) I am asked where my keys are. I look for them everywhere, until I shake my bag and hear the noise of keys inside.
- a. *Sie sind wohl in meiner Tasche.*
they are WOHL in my bag
'They're in my bag (I infer)'.
 - b. *Na-sa bag ko yata.*
in-LOC bag my yata
'They're in my bag (I infer)'.

As argued by Brugman & Macaulay (2015), inferential evidentials always belong to a closed group of grammatical words and the inference they express is not obtained through logic only. Considering that the data presented here show this behavior for both particles, and that they function correspondingly in each context for each language, we can safely conclude that both particles are inferential evidentials. In what follows, we will explore the differences between the two. *Yata* proves to be a modal evidential. As for *wohl*, the proposals of Zimmermann (2004), Repp (2006, 2009) and Egg and Mursell (2017) have in common that they all take *wohl* to be a SAO, and the proposals differ mainly in the way *wohl* is argued to operate on the speech act in detail, a discussion not relevant to the point we make in this paper. What matters is that these evidentials, despite their similar meaning contribution, need not necessarily belong to the same class of evidentials. The next section introduces the diagnostics used in the literature to distinguish the classes of modal from non-modal evidentials.

3. Evidentials: Modals vs SAOs

Frequently, the similarities and differences in their semantic-pragmatic behavior have been reflected in the literature on evidentials by distinguishing two types of them:⁶ modal evidentials and speech act operators. In this section, we first briefly outline some of the distinguishing features between modal and SAO evidentials,⁷ before we show that *yata* patterns with the former class and *wohl* with the latter.

In the spirit of Kratzer (1981; 1991), modal evidentials are said to behave like epistemic modals, operating on a propositional level, that is, they do not take scope over the proposition, quantifying over possible worlds that are compatible with some actual-world evidence (e.g. Izvorski 1997; Garrett 2001; Faller 2011; Matthewson et al. 2007; McCready & Ogata 2007). Forasmuch as they do not out-scope the proposition that contains them, they are interpreted within it. By its nature and in order to illustrate the contrast with SAOs, this allows for the speaker to directly question, doubt or disagree with the evidential component, thus passing the so-called ‘assent/dissent test’ (Faller 2002). SAO evidentials (Faller 2002 et

6. A third approach is that of evidentials as a kind of spatio-temporal operator, as proposed by Faller (2003) and Chung (2005, 2007), which we do not consider here for it is irrelevant to the discussion of *wohl* and *yata*, lacking any sort of spatial or temporal dimension.

7. For lack of space, we only relate the features of modal and SAO evidentials that are pertinent to the purposes of this work. For formal analyses of these two types of evidentials, we refer the interested reader to the references cited.

seq.; Davis, Potts & Speas 2007; Murray 2010), on the other hand, operate on an illocutionary level, and thus fail this test, for their meaning cannot be assented or dissented with. Additionally, SAOs further specify an illocutionary force. Basically, one can conceptualize every utterance as consisting of a propositional content and an illocutionary force scoping over it, which encodes information such as assertion, question, exclamation, and others, following seminal work by Austin (1962), Searle (1969), and Grice (1975). While modal evidentials operate as part of the propositional content of the utterance, SAOs have an effect on the illocutionary force of the clause and consequently take scope over the whole proposition. For example, Cuzco Quechuan reportative evidential *-si*, typically characterized as SAO, in a basic declarative sentence changes the illocutionary force of an assertion to that of *presentation* (Faller 2002 et seq.), that is, the speaker using the reportative is presenting the propositional content instead of asserting it.

The part of the utterance modified by the evidential, propositional content or illocutionary force, also correlates with other testable properties which can be used to determine the evidential's level of operation. Among the tests proposed to differentiate modal evidentials and SAOs (e.g. Lyons 1977; Faller 2002, 2007; Matthewson et al. 2007), Waldie et al. (2009) distinguish between those that are concerned with truth value, and those involving scope and embeddability. As we see in Table 1, not all of the tests are applicable to the particles. We must highlight, as Peterson (2010) already observes, that we should allow for many different factors related to the language that may be conditioning such non-applicability. For instance, as mentioned earlier, *yata* is unable to occur in interrogatives since it appears in complementary distribution with the speculative *kaya*. As for *wohl*, lexical restrictions seem to interfere with its ability to occur if *p* is known to be false. Despite lively debate on the reliability of these tests as applied across evidentials from different languages (e.g. Korotkova 2016),⁸ we show that *yata* patterns as a modal evidential (Tan 2016), and that *wohl* as a SAO.

8. Tests concerned with the felicity of the utterance when the prejacent is known to be true or false have been rejected as valid diagnostics in the literature. Matthewson et al. (2007) and von Stechow & Gillies (2010) refer to the necessity of distinguishing between strength of assertion and directness of evidence when the prejacent is known to be true, and thus involves factors preventing its applicability. As for known-to-be-false prejacent, AnderBois (2014) shows that, crosslinguistically, reportative evidentials can be felicitously used in such conditions due to pragmatic perspective shift, which may be independent from their being modal or SAO evidentials, thus restricting the validity of this test for distinctive purposes.

Table 1. Modal vs SAO evidentials tests and their application to *yata* and *wohl*

| | Modal | SAO | <i>yata</i> | <i>wohl</i> |
|----------------------------------------------|-------|-----|-------------|-------------|
| TESTS REGARDING TRUTH VALUE: | | | | |
| Felicitous if <i>p</i> is known to be false? | ✗ | ✓ | ✗ | n.a. |
| Pass assent/dissent test? | ✓ | ✗ | ✓ | ✗ |
| TESTS INVOLVING EMBEDDABILITY AND SCOPE | | | | |
| Semantically embeddable? | ✓ | ✗ | ✓ | (✓) |
| Able to scope outside interrogatives? | ✗ | ✓ | n.a. | ✓ |

3.1 *Yata* as a modal evidential

As mentioned earlier, *yata* cannot occur in interrogatives, thus the possibility of scoping out of interrogatives cannot be used as a diagnostic in this case. The factor conditioning this restriction is independent from *yata*'s status as a modal evidential and due to the existence of a dedicated evidential marker for questions, the particle *kaya*, that is, it is simply a lexical restriction.

In light of modal evidentials operating on a propositional level, they should pass the assent/dissent test, which states that if a given item can be disagreed with or questioned, it contributes to the truth conditions of the proposition (Faller 2002), which thus leads to a modal analysis. If not, this should make it a SAO. Matthewson et al. (2007: 32) provide the example in (10), where speaker B disagrees with the speaker's modal claim that *Jo must be the thief*, for it is not necessarily the case and other possibilities remain open. Likewise, *yata* is disagreed with by speaker B in (11), as they seem to know that speaker A could not possibly have inferred that *Pablo would come along*. They had not heard from him yet and therefore could not have any indirect evidence for *p*. Hence, *yata* passes the assent/dissent test.

- (10) a. *Jo must be the thief.*
 b. *That's not true. There are some other plausible suspects. Jo may be entirely innocent.*
- (11) a. *Sasama yata si Pablo.*
 will.come YATA NOM.PERS Pablo
 'Pablo will come along (I infer)'.
 b. *Hindi a. Hindi pa naman siya nagpaparinig, baka*
 not oh not still actually (s)he cause.to.be.heard maybe
umayaw pa 'yan.
 decline still that
 'Oh no. We haven't even actually heard from him yet, he might still decline'.

An additional test used to distinguish between modal evidentials and SAO evidentials is embedding. Only few languages are reported to allow for their evidentials to embed (Korotkova 2016), and it has been claimed that only modal evidentials, as part of the propositional content, can be semantically embedded (see data in 3.3). Further evidence of this is shown in (12), where *yata*, just like epistemic modals, is licensed only in so-called representational attitudes (Anand & Hacquard 2013). These include doxastic verbs (e.g. ‘think’), argumentation verbs (e.g. ‘say’), and semifactives (e.g. ‘realize’) and exclude desideratives (e.g. ‘want’) and directives (e.g. ‘demand’).

- (12) a. *Akala ko na kumain yata siya ng gulay.*
believe I that ate yata s/he GEN vegetables
‘I believe that s/he ate vegetables (I infer)’.
- b. *Sabi ni Pablo na sasama yata siya.*
say GEN.PERS Pablo that will.come yata he
‘Pablo says that (he infers) he will come’.
- c. *Naintindihan ng guro na wala yata-ng pasok*
realised GEN teacher that NON.EXIS yata-LNK class
‘The teacher realised that (she infers) there is no class’.
- d. **Gusto ni Pablo na bumalik yata sa Pilipinas.*
want GEN.PERS Pablo that return yata LOC Philippines
Int.: ‘Pablo wants to (I/he infer(s)) return to the Philippines’.
- e. **Nang-aangkin si Pablo na umulan yata.*
demand NOM.PERS Pablo that rain yata
Int.: ‘Pablo demands that it rains (I/he infer(s))’.

As stated earlier, it was argued that SAOs could not be embedded (Faller 2002) because they are not part of the propositional content. However, we show below that this dichotomy does not hold, since *yata*, as well as *wohl*, can both be embedded, the latter however only in a restricted set of contexts. To sum up, *yata* behaves like an epistemic modal and is therefore a modal evidential.

3.2 Wohl as a SAO

Several analyses of *wohl* in the literature propose a treatment of the DiP in terms of SAO. To argue that *wohl* should be treated as a SAO, it needs to be shown that, on the one hand, it is not part of the propositional content of the utterance, and on the other hand, that it does not behave like an expressive element, operating on a completely different level of meaning. In this argumentation, we follow Zimmermann (2004).

One reason to assume that *wohl* is not part of the expressed proposition is the observation that the DiP only takes full propositions as its argument, and thus, semantic binding across *wohl* and other DiPs is impossible.

- (13) *Jeder_x von den Arbeitern wurde entlassen, weil er_x wohl in
 everyone of the workers was fired because he WOHL in
 der KPD war.
 the KPD was

Int.: 'Every worker was fired because he was probably in the KPD'.

In (13), the variable introduced by the pronoun *er* in the embedded clause is bound by the quantifier *jeder* in the matrix clause and the presence of *wohl* in the embedded clause causes ungrammaticality. One way to account for this ungrammaticality is to assume that *wohl* needs to scope over full propositions. However, in (13), the proposition the DiP scopes over contains an open variable and thus does not count as a full. Scoping over propositions is a defining property of SAOs, which consequently suggests an analysis of *wohl* in terms of this category.

In addition to not being part of the proposition, it can also be shown that *wohl* does not behave like an expressive in that it has to be interpreted inside its host clause. Expressives, on the other hand, always take widest scope, referring to the speaker of the whole utterance.

- (14) a. Bush said that the damned Republicans deserve support.
 <Bush said the _ Republicans deserve support; Speaker does not like the Republicans>
- b. *Merkel hat gesagt dass alle Parteien wohl Unterstützung verdienen*
 Merkel has said that all parties WOHL support
 verdienen
 deserve
 ✗ <Merkel said that all parties _ deserve support; Speaker is unsure whether all parties deserve support>
 ✓ <Merkel said that all parties _ deserve support; Merkel is unsure whether all parties deserve support>

The important difference in (14) concerns the difference in attitude holder to which the expressive or DiP is related. In (14a), 'damn' as expressive indicates an attitude of the speaker of the whole utterance, meaning it is not Bush who dislikes the Republicans. Thus, the expressive takes widest scope, scoping out of its embedded clause. In (14b), such an interpretation is impossible for the DiP *wohl*. The particle in (14b) indicates an attitude of the speaker of the embedded clause, not the whole utterance, which in this case is Merkel. However, a note of caution concerning examples like (14) is in order. As an anonymous reviewer correctly points out, it

is easily possible to find examples in which an expressive does not scope out of its embedded context. Consider, for example, the sentence in (15), going back to Kratzer (1999), which clearly allows an interpretation relating the expressive *bastard* in the embedded clause to the matrix subject *my father* and not to the speaker of the whole utterance.

- (15) My father screamed he would never let me marry that bastard Webster.

Crucially, the possibility of relating an embedded expressive to the matrix subject as in (15) depends on the matrix predicate. The verb ‘scream’, as a manner-of-speaking verb imposes heavy restrictions on relations from the matrix into the embedded clause, it is for example not possible to extract *wh*-elements from complements embedded under such non-bridge verbs, as seen in (16b).⁹

- (16) a. Who did Mary say that Susan met?
b. *Who did Mary whisper that Susan met?

Note, however, that the matrix verb in (14) is not a manner-of-speaking verb and thus we do not expect any restrictions. This is borne out for the expressive in (14a) but the DiP is still unable to scope out of the complement clause. From that, we conclude that DiPs do indeed behave differently from expressives when it comes to scoping possibilities out of embedded clauses.

Summing up the preceding discussion, *wohl* is not interpreted as part of the proposition, nor does it behave like an expressive. Therefore, the question arises how to best analyse the DiP’s contribution to the clause. Following the literature presented above, we will assume that *wohl* works as a SAO, modifying the speech act of the clause it occurs in. Something comparable has also been proposed by Jacobs (1984) for *ja*, another German DiP.

Of course, the question of how this modification is achieved needs to be addressed. DiPs are mostly restricted to a medial position in the clause. In approaches that assume an encoding of the speech act in the clause (Rizzi 1997), the speech act or illocutionary force is usually hosted in the sentence initial periphery. We follow Bayer & Obenauer (2011), Coniglio & Zegrean (2012) and Egg & Mursell (2017) in assuming that this connection is based on specific features, more specifically information structural features like focus. Since this discussion is far beyond the scope of this paper, the interested reader is referred to the literature cited above.

Turning to the tests used to distinguish modal evidentials from SAO evidentials, it can be shown that they also suggest a treatment of *wohl* as SAO. Similarly

9. This is just a superficial discussion of the complex phenomenon of contextually shifted indexicals which we cannot go into here. For detailed discussion of the problem, cf. especially Kaplan (1989), Schlenker (2003), and Anand (2006).

to *yata*, however, not all tests are applicable. As already discussed with respect to Example (5) above, *wohl* cannot be used when the utterance contains a fact that is generally known. This does not only hold for propositions that are generally known to be true but also those that are generally known to be false. Thus, due to a lexical restriction of *wohl*, this test is not applicable to determine *wohl*'s level of operation.

Turning to the assent/dissent test, *wohl* behaves as expected of a SAO in that the meaning contribution cannot be assented or dissented with (17b). It is of course possible to dissent with the whole proposition, (17c), but importantly it is impossible to dispute the inferential contribution of the DiP.

- (17) a. *Peter kommt wohl noch.*
 Peter comes WOHL still
 'Peter is still coming, I assume.'
- b. **Nein, du hast keinen Grund das zu vermuten.*
 No, you have no reason that to assume
 'No, you don't have a reason to assume that.'
- c. *Nein, wird er nicht, wir haben nichts von ihm gehört.*
 No, will he not, we have nothing from him heard
 'No he won't, we haven't heard anything from him.'

Looking at *wohl* in questions, *wohl* scopes over illocutionary operators like the question operator, which is used by Zimmermann (2004) to argue in favour of a SAO analysis of *wohl*. Following Hamblin (1973) and Karttunen (1977), we assume that the meaning of a question is the set of its possible answers. This set is also called the proto-question and the illocutionary question operator ? turns this set into a proper question.

- (18) a. Is it windy?
 b. $p =$ it is windy
 c. $\{p, \neg p\} =$ {it is windy, it is not windy} (proto-question)
 d. $?\{p, \neg p\} =$?{it is windy, it is not windy}
 \approx Tell me what is the case: it is windy or it is not windy.

If *wohl* was part of the proposition, questions containing it should behave similarly to (18). However, as (19) shows, this is not the case.

- (19) *Ist Hein wohl auf See?*
 is Hein WOHL at sea
 'Is Hein at sea (what do you assume)?'
 ✗ $?\{\text{ASSUME}'(\text{Hearer}, \text{H. is at sea}), \neg\text{ASSUME}'(\text{Hearer}, \text{H. is at sea})\}$
 'Do you assume that H. is at sea or do you not assume that H. is at sea?'
 ✓ $?\text{ASSUME}'(\text{Hearer}, \{\text{H. is at sea}, \text{H. is not at sea}\})$
 'Tell me your assumption about whether or not H. is at sea.'

As shown in (19), the DiP cannot be interpreted as part of the proto-question. Thus, the question is not about whether the hearer assumes something or not, but about the assumption whether or not something is the case. Consequently, questions with *wohl* form the same proto-question as those without it, and the DiP then scopes over the question operator.

Similarly to *yata*, we will discuss the embeddability of *wohl* separately in the next section. Looking at the literature, we expect *wohl*, in contrast to *yata*, to be impossible in embedded contexts since it depends on the presence of independent illocutionary force. However, we will show that there is a well-defined set of contexts in which embedding *wohl* is possible. As expected, the contexts in which *wohl* can be embedded are those that have independently been argued to contain illocutionary force.

3.3 Embeddability test restrictions

In this section, we focus on the embeddability test since it appears to contradict the analysis of *wohl* as a SAO which predicts unembeddability. We show that this test requires an addendum, namely, that SAOs, such as *wohl*, can be embedded in a well-defined subset of environments in which modal evidentials, such as *yata*, may occur (contra previous claims in Faller 2002; Matthewson et al. 2007; although see Faller 2014 for embedding Cuzco Quechua reportative *-si* under say verbs). In general, embedding *wohl* is possible in contexts that have been described as embedded root contexts (Heycock 2005), that is, embedded environments that allow transformations usually restricted to main clauses. In general, root contexts differ from non-root contexts in the applicability of certain operations, for example Germanic V2. Initially, root-contexts were equivalent to main clauses, while all embedded clauses were treated as non-root-contexts. However, over time it emerged that certain embedded contexts allowed for the same operations as main clauses, and those embedded contexts were consequently termed embedded root-contexts. In particular, embedded root contexts are complements of non-factive predicates, certain (peripheral) adverbial clauses,¹⁰ and non-restrictive relative clauses.

10. Schenner & Sode (2014) already put forward the possibility of embedding *wohl* in causal clauses for the same reason. We further check its embeddability in the rest of embedded root-contexts.

3.3.1 *Factive vs non-factive predicates*

Factive predicates presuppose the truth of their complement. Both evidentials can be embedded under non-factive predicates (20, 21), but, unlike *yata* (23), *wohl* is disallowed under factive predicates, as shown in (22). Importantly, clauses embedded under non-factive predicates are a typical example of embedded root clauses, while clauses embedded under factive verbs cannot be considered to have root status.¹¹

- (20) a. *Peter hat gesagt, dass die NPD wohl verboten wird.*
 Peter has said that the NPD WOHL banned will
 ‘Peter said that the NPD will be banned (he infers)’.
- b. *Maria glaubt, dass Peter wohl zu Hause geblieben ist.*
 Maria believes that Peter WOHL at home stayed has
 ‘Maria believes that Peter stayed at home (she infers)’.
- (21) a. *Sabi ni Pedro na hindi yata darating ang hepe*
 say GEN.PERS Pedro that not YATA will.come NOM boss
ngayon.
 yesterday
 ‘Pedro says that our boss will not come today (he infers)’.
- b. *Akala ni Pedro na may kumatok yata sa pinto.*
 believe GEN.PERS Pedro that EXIS knocked YATA LOC door
 ‘Pedro believes that someone knocked on the door (he infers)’.
- (22) a. **Er leugnete, dass er die Zeugin wohl unter Druck*
 he denied that he the witness WOHL under pressure
gesetzt habe.
 put had
 Int.: ‘He denied that he (he/I infer(s)) put pressure on the witness’.
- b. **Es ist schön dass Maria wohl morgen kommt.*
 it is nice that Maria WOHL tomorrow comes
 Int.: ‘It’s nice that Maria (I infer) arrives tomorrow’.

11. One class of factive verbs seems to allow *wohl* in certain contexts, namely factive attitude predicates.

- (i) *Franz ist eingefallen, dass Klaus wohl allergisch auf Nüsse ist.*
 F. is remembered that K. wohl allergic on Nuts is
 ‘Franz remembered that Klaus is allergic to nuts (but I, the speaker, am not sure about this fact)’.

Importantly, in these instances, the DiP does not modify the illocutionary force of the embedded clause but the matrix force and thus relates to the speaker of the whole utterance. Thus, the DiP in these instances seems to behave similarly to an expressive. We leave an in-depth investigation of this particular phenomenon to further research.

- (23) a. *Naalala ni kuya na maggagawa yata si*
 remembered GEN.PERS brother that will.do YATA NOM.PERS
nanay ng handaan.
 mother GEN party
 ‘My brother remembered that (as he inferred) mom is doing a party’.
- b. *Totoo na hindi yata sasama si tatay sa handaan.*
 be.true that not YATA will.come NOM.PERS father LOC party
 ‘It’s true that (as I infer) my father is not coming to the party’.

3.3.2 Relative clauses: Restrictive vs non-restrictive

As for relative clauses, *wohl* can only occur in non-restrictive clauses (24), whereas *yata* can also appear in restrictive relative clauses (25). Thus, relative clauses that are ambiguous between a restrictive and a non-restrictive reading are disambiguated by *wohl*.

- (24) *Die Deutschen, die wohl viel Bier trinken, gucken gerne Fussball.*
 the Germans who WOHL much beer drink watch gladly football
 ✓ ‘The Germans, who (generally) like to drink much beer (I infer), like to watch football’.
 ✗ ‘Those Germans who like to drink much beer (I infer) like to watch football’.
- (25) a. *Naiwan ko sa bahay ang susi na nakakabukas yata ng*
 left I LOC house NOM key that can.open YATA GEN
garahe.
 garage
 ‘I left at home the key that (I infer) can open the garage’.
- b. *Kinain ni Juan ang mamon, na ginawa yata ni*
 ate GEN.PERS Juan NOM mamon that made YATA GEN.PERS
lola.
 grandma
 ‘Juan ate the muffin, which (I infer) grandma made’.

Similarly to the distinction between clauses embedded under factive and non-factive predicates, non-restrictive relative clauses are usually considered to be embedded root clauses, in contrast to restrictive relative clauses.

3.3.3 Adverbial clauses

Lastly, *wohl* is also only possible in a subset of adverbial clauses while *yata* does not have such restriction (27). For instance, *wohl* is fine in concessive adverbial clauses (26a), but banned from locative ones (26b).

- (26) a. *Er bestand die Klausur nicht, obwohl er wohl recht intelligent ist.*
 he passed the exam not even.though he WOHL rather intelligent is
 intelligent is
 ‘He didn’t pass the exam, even though (I infer) that he is rather intelligent’.
- b. **Er lebt, wo das Stadium wohl gebaut wurde.*
 he lives where the stadium WOHL built was
 Int.: ‘He lives where (I infer) the stadium was built’.
- (27) a. *Pumasok siya sa trabaho kahit may sakit yata siya.*
 entered (s)he LOC work even.though EXIS illness YATA (s)he
 ‘He went to work even though (I infer) he is sick’.
- b. *Lumindol sa Mindanao, kung saan yata nakatira si tito.*
 there.be.earthquake LOC Mindanao if where YATA lives
 si tito.
 NOM.PERS uncle
 ‘There was an earthquake in Mindanao, where (I infer) grandpa lives’.

The SAO *wohl* is possible in those adverbial clauses that have been analysed as peripheral adverbial clauses by Haegeman (2006 et seq.). Peripheral adverbial clauses, in contrast to central ones, have been claimed to be embedded root-clauses, considering that these occur in a layer where association with speaker deixis is available (Haegeman 2006: 1667). Hence, this allows for SAO evidentials like *wohl* to occur in them.

3.4 Force as licenser for *wohl*

Considering the data in the last section, the question arises, why SAOs should only be embeddable in a restricted set of sentences. One promising line of analysis concerning DiPs argues that they modify illocutionary force, which we equate with the speech act (but see Heim et al. 2014 for further refinements), encoded in a position scoping above everything else in the clause. Consequently, we expect that if a clause does not contain independent illocutionary force, it should not be able to host a SAO. In other words, following Haegeman (2006, 2012) as well as Coniglio (2008), we claim that SAOs are embeddable in those embedded clauses that contain independent accessible illocutionary force, so called root-like clauses or embedded root-contexts. The simplified left peripheries of root- and non-root clauses are shown in (28), with most of the functional structure being absent in non-root contexts.

- (28) a. Root clauses: [(Sub) [_{ForceP} [_{TopP*} [_{FocP} [_{ModP*} [_{FinP} ...]]]]]]]]
 b. Non-root clauses: [(Sub) [_{ModP*} [_{FinP} ...]]]]

As expected, root clauses or root-like embedded clauses according to the authors are complements of non-factive predicates, non-restrictive relative clauses, and peripheral (but not central) adverbial clauses.

Several analyses of how to formalise the connection between illocutionary force and SAO, are possible. However, since all analyses share the dependence of *wohl* on independent illocutionary force, which enables us contrasting it with *yata*, and since this is the main argument of this chapter, we will not discuss specific analyses in more detail at this point and refer the reader to the above mentioned works.

4. Conclusions and further issues

We have presented a view of the German *wohl* and the Tagalog *yata* as evidentials conveying both information source and degree of certainty, by comparing and examining their occurrence in different contexts that target either uncertainty alone or both uncertainty and information source. Their infelicity in the former is accounted for in an evidential treatment of both particles. With such similar meaning contributions, we have argued further that they should be considered inferential evidentials, since the contexts provided here show that they require for the speaker to infer their claim based on certain indirect evidence, via observable results or reasoning. Upon portraying the relevant properties of each particle and applying the pertinent standard tests for their analysis, we have determined that *wohl* should be considered a SAO type of evidential and *yata* a modal evidential. This claim is further justified by systematically contrasting the environments in which each particle can be embedded. While modals, such as *yata*, may occur with all types of embedding predicates, SAOs such as *wohl* cannot, as they are restricted to root-like clauses. Thus, we have also shown that testing their embeddability with different types of predicates provides a reliable distinction between modal and non-modal analyses.

Certain issues require some further research. Many more aspects of the syntax, semantics, and pragmatics of each individual particle may be investigated. Furthermore, it would be interesting to analyze the scope of the evidentials, as they are taking scope only over the embedded clauses in most cases. Throughout, we have mostly provided the particles' translation in parenthesis, in which we show how they are interpreted within the embedded clause, since the evidential origo is shifted to the subject of the matrix clause. However, this is of course limited (see footnote 7), a more detailed investigation into a wider variety of contexts appears to be necessary. Furthermore, we briefly mentioned the fact the Tagalog uses a different particle, the speculative marker *kaya*, to express a similar meaning to *wohl* in questions. The contrasts between these and other such particles may further shed light onto the relation between DiPs and evidentials.

Acknowledgements

This research is supported by Spanish Ministry of Economy and Competitiveness grants FFI2014-56968-C4-3-P and BES2013-062583 and by the DFG in the form of the SFB 632 A8. Thanks to the participants at TAM-E: Comparative, Cognitive, Theoretical & Applied Perspectives conference at Paris Diderot University (November 17–18, 2016), as well as the Syntax group at Frankfurt University. Remaining errors and shortcomings are ours.

References

- Abraham, Werner & Leiss, Elisabeth (eds). 2009. *Modalität. Epistemik und Evidentialität bei Modalverb, Adverb, Modalpartikel und Modus*. Tübingen: Stauffenburg.
- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: OUP.
- Aikhenvald, Alexandra Y. (ed.). 2018. *The Oxford Handbook of Evidentiality*. Oxford: OUP.
- Anand, Pranav. 2006. *De De Se*. PhD dissertation, MIT.
- Anand, Pranav & Hacquard, Valentine. 2013. Epistemics and attitudes. *Semantics and Pragmatics* 6(8): 1–59.
- AnderBois, Scott. 2014. On the exceptional status of reportative evidentials. In *Proceedings of Semantics and Linguistic Theory*, Todd Snider, Sarah D'Antonio & Mia Weigand (eds), 24: 234–254. doi:10.3765/salt.v24io.2424
- Anderson, Stephen R. 2008. Second position clitics in Tagalog. *The Nature of the Word*, Kristin Hanson & Sharon Inkelas (eds), 549–466.
- Austin, John L. 1962. *How to Do Things with Words*. Oxford: Clarendon.
- Barbet, Cécile. 2012. Devoir et pouvoir, des marqueurs modaux ou évidentiels? *Langue Française* 173: 49–63. doi:10.3917/lf.173.0049
- Bayer, Josef & Obenauer, Hans-Georg. 2011. Discourse particles, clause structure, and question types. *The Linguistic Review* 28, 449–491. doi:10.1515/tlir.2011.013
- Brugman, Claudia & Macaulay, Monica. 2015. Characterizing evidentiality. *Linguistic Typology* 19(2): 201–238. doi:10.1515/lingty-2015-0007
- Cardinaletti, Anna & Starke, Michal. 1999. The typology of structural deficiency. In *Clitics in the Languages of Europe*, Henk van Riemsdijk (ed.), 145–234. Berlin: Mouton de Gruyter.
- Chafe, Williams & Nichols, Johanna (eds). 1986. *Evidentiality: The Linguistic Coding of Epistemology*. Norwood NJ: Ablex.
- Chung, Kyung-Sook. 2005. *Space in Tense: The Interaction of Tense, Aspect, Evidentiality and Speech Act in Korean*. PhD dissertation, Simon Fraser University.
- Chung, Kyung-Sook. 2007. Spatial deictic tense and evidentials in Korean. *Natural Language Semantics* 15: 187–219. doi:10.1007/s11050-007-9017-8
- Coniglio, Marco. 2008. *Die Syntax der Deutschen Modalpartikeln: Ihre Distribution und Lizenzierung in Haupt- und Nebensätzen*. PhD dissertation, Università Ca'foscari Venezia.
- Coniglio, Marco & Zegrean, Iulia. 2012. Splitting up force: Evidence from discourse particles. In *Main Clause Phenomena. New horizons* [Linguistik Aktuell/Linguistics Today 190], Lobke Aelbrecht, Liliane Haegeman & Rachel Nye (eds), 229–255. Amsterdam: John Benjamins. doi:10.1075/la.190.10con

- Cornillie, Bert. 2009. Evidentiality and epistemic modality. On the close relationship between two different categories. *Functions of Language* 16(1): 44–62. doi:10.1075/fol.16.1.04cor
- Davis, Christopher, Potts, Christopher & Spears, Margaret. 2007. The pragmatic values of evidential sentences. *Proceedings of SALT*, Tova Friedman and Masayuki Gibson (eds), 17, 71–88. Ithaca NY: CLC Publications. doi:10.3765/salt.v17i0.2966
- Dendale, Patrick. 1994. Devoir: Marqueur modal ou évidentiel? *Langue Française* 102: 24–40. doi:10.3406/lfr.1994.5712
- Dendale, Patrick. 2001. Le futur conjectural versus devoir épistémique: Différences de valeur et restrictions d'emploi. *Le Français Moderne* 69: 1–20.
- Dendale, Patrick & Tasmowski, Liliane. 2001. Introduction: Evidentiality and related notions. *Journal of Pragmatics* 33: 339–348. doi:10.1016/S0378-2166(00)00005-9
- Diewald, Gabriele. 2006. Discourse particles and modal particles as grammatical elements. In *Approaches to Discourse Particles*, Kerstin Fischer (ed.), 403–425. Amsterdam: Elsevier.
- Eckardt, Regine. 2017. Engaging questions: Wohl in German V_{end} -questions. Paper presented at the Non-At-Issue Meaning and Information Structure Workshop. University in Oslo, May 8–10.
- Egg, Markus & Mursell, Johannes. 2017. The syntax and semantics of discourse particles. In *Discourse Particles. Formal Approaches to Their Syntax and Semantics*, Josef Bayer & Volker Struckmeier (eds), 15–48. Berlin: Mouton de Gruyter.
- Faller, Martina. 2002. *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. PhD dissertation, Stanford University.
- Faller, Martina. 2003. Propositional- and illocutionary-level evidentiality in Cuzco Quechua. *Proceedings of SULA*, Jan Anderssen, Paula Menéndez-Benito & Adam Werle (eds), 2, 19–33. Amherst MA: GLSA.
- Faller, Martina. 2007. *Evidentiality above and below speech acts*. Ms. University of Manchester.
- Faller, Martina. 2011. A possible world semantics for Cuzco Quechua evidentials. In *Proceedings of Semantics and Linguistic Theory 20*, Li Nan & David Lutz (eds), 660–683. Ithaca, NY: CLC Publications.
- Faller, Martina. 2014. Do speech act evidentials embed after all? Paper presented at the Workshop on the Semantics of Embedding and the Syntax of the Left Periphery, LAGB, Oxford, September.
- von Fintel, Kai & Gillies, Anthony S. 2010. Must... stay... strong!. *Natural Language Semantics* 18: 351–383. doi:10.1007/s11050-010-9058-2
- Garrett, Edward. 2001. *Evidentiality and Assertion in Tibetan*. PhD dissertation, University of California.
- Givón, Talmy. 2001. *Syntax: An Introduction*, Vol. 1. Amsterdam: John Benjamins.
- Grice, H. Paul. 1975. Logic and conversation. In *The Logic of Grammar*, Donald Davidson & Gilbert Harman (eds), 64–75. Encino CA: Dickenson.
- Gutzmann, Daniel. 2015. *Use-Conditional Meaning: Studies in Multidimensional Semantics*. Oxford: OUP. doi:10.1093/acprof:oso/9780198723820.001.0001
- de Haan, Ferdinand. 1999. Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics*: 18 91–105.
- de Haan, Ferdinand. 2001. The relation between modality and evidentiality. *Linguistische Berichte Sonderheft 9*: 201–216.
- Haegeman, Liliane. 2006. Conditionals, factives and the left periphery. *Lingua* 116: 1651–1669. doi:10.1016/j.lingua.2005.03.014

- Haegeman, Liliane. 2012. *Adverbial Clauses, Main Clause Phenomena, and Composition of the Left Periphery*. Oxford: OUP. doi:10.1093/acprof:oso/9780199858774.001.0001
- Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10: 41–53.
- Haumann, Dagmar & Letnes, Ole. 2012. German *wohl*: An evidential? In *Covert Patterns of Modality*, Werner Abraham & Elisabeth Leiss (eds), 202–238. Newcastle upon Tyne: Cambridge Scholars.
- Heim, Johannes, Keupdjio, Hermann, Wai-Man Lam, Zoe, Osa-Gómez, Adriana & Wiltschko, Martina. 2014. How to do things with particles. In *Proceedings of the Annual Conference of the Canadian Linguistic Association*, Laura Teddiman (ed.), 1–15. Toronto: Canadian Linguistic Association.
- Helbig, Gerhard. 1988. *Lexikon Deutscher Partikeln*. Leipzig: Enzyklopädie.
- Hentschel, Elke. 1986. *Funktion und Geschichte Deutscher Partikeln*. Tübingen: Niemeyer. doi:10.1515/9783111371221
- Heycock, Caroline. 2005. Embedded root phenomena. In *The Blackwell Companion to Syntax*, Martin Everaert & Henk van Riemsdijk (eds), 179–209. Oxford: Blackwell.
- Izvorski, Roumyana. 1997. The present perfect as an epistemic modal. In *Proceedings of the 7th Semantics and Linguistic Theory Conference, held March 21–23 1997 at Stanford University*, Aaron Lawson (ed.), 222–239. <<https://journals.linguisticsociety.org/proceedings/index.php/SALT/issue/view/102>>
- Jacobs, Joachim. 1984. Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte* 91: 25–58.
- Jacobsen, William H. 1986. The heterogeneity of evidentials in Makah. In *Evidentiality: The Linguistic Coding of Epistemology*, William Chafe & John Nichols (eds), 3–28. Norwood NJ: Ablex.
- Kaplan, David. 1989. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In *Themes from Kaplan*, Joseph Almog, John Perry & Howard Wettstein (eds), 481–563. Oxford: OUP.
- Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1: 1–44. doi:10.1007/BF00351935
- Kaufman, Daniel. 2010. *The Morphosyntax of Tagalog Clitics: A Typologically Driven Approach*. PhD dissertation, Cornell University.
- van der Klok, Jozina. 2014. *On the use of questionnaires in semantic fieldwork: A case study in modality*. Ms, University of British Columbia.
- Korotkova, Natalia. 2016. *Heterogeneity and Uniformity in the Evidential Domain*. PhD dissertation, University of California at Los Angeles.
- Kratzer, Angelika. 1981. The notional category of modality. In *Words, Worlds and Contexts*, Hans J. Eikmeyer & Hannes Rieser (eds), 38–74. Berlin: Mouton de Gruyter.
- Kratzer, Angelika. 1991. Modality. *Semantics: An International Handbook of Contemporary Research*, Dieter Wunderlich & Arnim von Stechow (eds), 639–650. Berlin: Mouton de Gruyter.
- Kratzer, Angelika. 1999. Beyond *ouch* and *oops*. Paper presented at the Cornell Conference on Theories of Context Dependency. Ithaca NY, Cornell University, March 26.
- Kroeger, Paul. 1998. Clitics and clause structure in Tagalog. In *Pagtanaw: Essays on Language in Honor of Teodoro A. Llamzon*, 53–72. Manila: Linguistic Society of the Philippines.
- Kroeger, Paul. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford CA: CSLI.
- Lyons, John. 1977. *Semantics*, Vol. 2. Cambridge: CUP.

- Matthewson, Lisa. 2012. When fieldwork meets theory: Evidence about evidentials. In *Empirical Approaches to Linguistic Theory: Studies of Meaning and Structure*, Britta Stolterfoht & Sam Featherston (eds), 85–114. Berlin: Mouton de Gruyter.
- Matthewson, Lisa. 2015. Evidential restrictions on epistemic modals. In *Epistemic Indefinites*, Luis Alonso-Ovalle & Paula Menéndez-Benito (eds), 141–160. Oxford: OUP.
doi:10.1093/acprof:oso/9780199665297.003.0007
- Matthewson, Lisa, Rullmann, Hotze & Davis, Henry. 2007. Evidentials as epistemic modals: Evidence from St'át'imcets. *Linguistic Variation Yearbook* 7: 201–254. doi:10.1075/livy.7.07mat
- McCready, Eric & Ogata, Norry. 2007. Evidentiality, modality and probability. *Linguistics and Philosophy* 30(2): 147–206. doi:10.1007/s10988-007-9017-7
- McCready, Eric. 2015a. Evidentiality. Lecture given at the 27th European Summer School in Logic, Language and Information, Universitat Pompeu Fabra, August.
- McCready, Eric. 2015b. *Reliability in Pragmatics*. Oxford: OUP.
- Modicom, Pierre-Yves. 2012. Shared knowledge and epistemic reductionism: Covert semantics of German modal particles. In *Covert Patterns of Modality*, Werner Abraham & Elisabeth Leiss (eds), 202–238. Newcastle upon Tyne: Cambridge Scholars.
- Murray, Sarah. 2010. *Evidentiality and the Structure of Speech Acts*. PhD dissertation, Rutgers University.
- Nuyts, Jan. 2006. Modality: Overview and linguistic issues. In *The Expression of Modality*, William Frawley (ed.), 1–26. Berlin: Mouton de Gruyter.
- Palmer, Frank. 1986. *Mood and Modality*. Cambridge: CUP.
- Peterson, Tyler. 2010. *Epistemic Modality and Evidentiality in Gitksan at the Semantics-Pragmatics Interface*. PhD dissertation, University of British Columbia.
- Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(2): 165–198.
doi:10.1515/TL.2007.011
- Repp, Sophie. 2009. *Negation in Gapping*. Oxford: OUP.
doi:10.1093/acprof:oso/9780199543601.001.0001
- Repp, Sophie. 2006. ¬(A&B). Gapping, negation and speech act operators. In *Special issue Ellipsis*, Jennifer Spnader & Petra Hendricks (eds). *Research on Language and Computation* 4(4): 397–423.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar*, Liliane Haegeman (ed.), 281–337. Dordrecht: Kluwer. doi:10.1007/978-94-011-5420-8_7
- Rooryck, Johan. 2001. Evidentiality, Part I. *Glott International* 5: 125–133.
- de Saussure, Louis. 2014. Inferential evidentiality with epistemic modality. A pragmatic stance. Paper presented at *Chronos 11*, Scuola Normale Superiore, Pisa, 16–18 June.
- Schachter, Paul. 1973. Constraints on clitic order in Tagalog. In *Parangal kay Cecilio Lopez*, Andrew B. Gonzalez (ed.), 214–231. Manila: Linguistic Society of the Philippines.
- Schachter, Paul & Otones, Fe. 1972. *Tagalog Reference Grammar*. Berkeley CA: University of California Press.
- Schenner, Mathias & Sode, Frank. 2014. Modal particles in causal clauses. The case of German *weil wohl*. In *Modes of Modality* [Studies in Language Companion Series 149], Elisabeth Leiss & Abraham Werner (eds), 291–314. Amsterdam: John Benjamins. doi:10.1075/slcs.149.10sch
- Scherf, Nathalie. 2017. The syntax of Swedish modal particles. In *Discourse Particles. Formal Approaches to Their Syntax and Semantics*, Josef Bayer & Volker Struckmeier (eds), 78–99. Berlin: Mouton de Gruyter.

- Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and Philosophy* 26(1): 29–120.
doi:10.1023/A:1022225203544
- SEALang Library Tagalog Text Corpus. <<http://sealang.net/tagalog/corpus.htm>> (10 November 2016).
- Searle, John R. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: CUP.
doi:10.1017/CBO9781139173438
- Simons, Mandy. 2006. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117: 1034–1056. doi:10.1016/j.lingua.2006.05.006
- Speas, Margaret & Tenny, Carol. 2003. Configurational properties of point of view roles. In *Asymmetry in Grammar*, Vol. 1: *Syntax and Semantics* [Linguistik Aktuell/Linguistics Today 57], Anna-Maria DiSciullo (ed.), 315–343. Amsterdam: John Benjamins.
doi:10.1075/la.57.15spe
- Stalnaker, Robert. 2002. Common Ground. *Linguistics and Philosophy* 25(5): 701–721.
doi:10.1023/A:1020867916902
- Tan, Jennifer. 2016. On the modality of Tagalog evidentials. In *Proceedings of the 23rd Meeting of the Austronesian Formal Linguistics Association*, Hiroki Nomoto, Takuya Miyauchi & Asako Shiohara (eds), 273–287. Canberra: Asia-Pacific Linguistics.
- Thurmair, Maria. 1989. *Modalpartikeln und ihre Kombinationen*. Tübingen: Niemeyer.
doi:10.1515/9783111354569
- Tonhauer, Judith & Matthewson, Lisa. 2015. *Empirical evidence in research on meaning*. Ms. <<http://ling.auf.net/lingbuzz/002595>>
- Vanderveken, Daniel. 1990. *Meaning and Speech Acts*. Cambridge: CUP.
- Waldie, Ryan, Peterson, Tyler, Rullmann, Hotze & Mackie, Scott. 2009. Evidentials as epistemic modals or speech act operators: testing the tests. Paper presented at The 14th Workshop on Structure and Constituency in Languages of the Americas (WSCLA), Purdue University.
- Willett, Thomas. 1988. A crosslinguistic survey of the grammaticization of evidentiality. *Studies in Language* 12: 51–97. doi:10.1075/sl.12.1.04wil
- Zeevat, Henk. 2000. Discourse particles as speech act markers. *LDV-Forum* 17: 74–91.
- Zimmermann, Malte. 2004. Zum ‘wohl’: Diskurspartikeln als Satztypmodifikatoren. *Linguistische Berichte* 199: 253–286.

Questions as indirect speech acts in surprise contexts

Agnès Celle

Université Paris Diderot and University of Colorado

This chapter offers an analysis of two types of interrogatives used as indirect speech acts in surprise contexts in English – unresolvable questions and rhetorical questions. The function of these questions is not to request information that is unknown to the speaker. It is argued that surprise-induced unresolvable questions are expressive speech acts devoid of epistemic goals. Surprise-induced rhetorical questions are shown not to suggest an obvious answer, but to request a commitment update from the addressee. Adopting a schema-theoretic approach to surprise, it is shown that unresolvable questions and rhetorical questions can express mirativity, the former at the initial stage of the cognitive processing of unexpectedness, the latter at the last stage.

Keywords: rhetorical questions, conjectural questions, unresolvable questions, commitment, mirativity, surprise, expressivity

1. Introduction

This chapter offers an analysis of interrogative structures used as indirect speech acts in surprise contexts in English. Refining Littell et al.'s (2010) typology, I distinguish two different types of interrogative structures: those that are mapped on the default interpretation of interrogatives, that is, requests for information (analysed in Celle et al. *forthc.*), and those that are linked to other speech acts. The latter correspond to indirect speech acts and include unresolvable questions and rhetorical questions. These are the object of the present chapter.

In recent years, there has been a flurry of research into conjectural questions in connection either with the conditional, the epistemic future and the subjunctive in Romance languages (Diller 1977; Haillet 2001; Celle 2007; Rocci 2007; Dendale 2010; Bourova & Dendale 2013; Azzopardi & Bres 2014) or with evidentials in languages that grammatically encode evidentiality (Littell et al. 2010; San Roque et al. 2017). In Romance languages, conjectural questions with the conditional,

the epistemic future or the subjunctive are reported to put forward an inference that the addressee is asked to evaluate. In Amerindian languages, the insertion of an evidential into a question seems to give rise to a different meaning. Littell et al. (2010: 92) claim that in three Amerindian languages with an evidential system, “the insertion of a conjectural / inferential into a question creates a non-interrogative utterance, roughly translatable using ‘I wonder’”. In this chapter, it is argued that emotive modifiers in English cancel the interrogative force of a question in a similar way to those evidentials. Conjectural questions in English are to be understood as expressive questions expressing wondering and uncertainty. However, the label ‘conjectural’ may be misleading in this case, as these questions do not form a conjecture, but rather implicate that it is impossible to resolve the question. Therefore, I propose to label these questions **unresolvable** rather than **conjectural**.

Unresolvable questions and rhetorical questions do not constitute requests for information. In the surprise contexts under scrutiny, unresolvable questions function as outcome-related and speaker-oriented utterances expressing wonder and disbelief. In English, their expressive function is marked by interjections, emotive modifiers and deictic items. Rhetorical questions stand as argumentative tools questioning some prior surprising discourse entity or extralinguistic event. Rhetorical questions in surprise contexts highlight the connection between surprise and negatively-valenced emotions such as anger and disappointment.

The aim of this chapter is to determine the function of those questions that do not request an answer in a surprise context. What do they tell us about the speaker-addressee relationship? What is the relation between surprise and questions used as indirect speech acts? Are these questions the linguistic expression of mirativity, and how do they relate to the cognitive integration of unexpected new information? Section 2 presents the data. Section 3 is devoted to surprise-induced unresolvable questions and Section 4 to surprise-induced rhetorical questions.

2. The data

This study is part of a large-scale project on surprise.¹ It presents an analysis of verbal reactions to surprising situations in the scripts of three movies drawn from the OpenSubtitle corpus (*Ed Wood*, *War of the Worlds*, *Dr. Strangelove*).² All surprising

1. This study is part of the Emphiline project (ANR-11-EMCO-0005), a project funded by the National Research Agency: “la surprise au sein de la spontanéité des émotions: un vecteur de cognition élargie”.

2. These movies provide a wealth of surprising episodes. The reasons for this choice are spelled out in more detail in Celle et al. (forthc.).

episodes were coded using the annotation tool Glozz,³ based on the same annotation scheme as Celle et al. (forthc.). The data used are enacted data. I am aware that this type of data may bias the expression of emotions, actors being prone to overemphasise some cues (Scherer et al. 2011: 409) as they relive an emotional experience of their own.⁴ However, enacted data allow recognizing emotions in a reliable way.

First, stage directions from the movie scripts can provide important environmental information about the context and the experiencer's emotional state. Second, emotions in movies can be detected and identified through patterns of observable vocal, facial and bodily cues. On the basis of experimentally-induced surprise reactions, Reisenzein (2000: 29) stresses that surprise faces most frequently display only one of the facial components associated with surprise: eyebrow raising, eye widening, or mouth opening, while two- or three-component displays are less frequent. He further points out that this finding is in keeping with Carroll and Russell's (1997) enacted data based on the facial displays of surprise shown by movie actors. Even if the present chapter does not aim to analyze intonation and gestures, those parameters were taken into account in our annotation scheme and facilitated emotion recognition. The semasiological perspective adopted is thus combined with an onomasiological approach, that is, only interrogatives occurring in surprise contexts were considered.

Interrogative clauses used as indirect speech acts are questions that do not request an answer from the addressee, although they may call for some response from the addressee. They amount to 13% of all interrogatives in our sample (26 utterances out of a total of 146 interrogatives). These interrogative clauses are subdivided into rhetorical questions ($n = 12$), unresolvable questions ($n = 5$), and clarification requests ($n = 9$).⁵ Like the interrogative clauses used as direct speech acts examined in Celle et al. (forthc.), the interrogative clauses found in our sample are triggered by some surprising event.⁶ Unlike their counterparts used as direct speech acts, however, they have no force of inquiry. The connection between surprise and interrogative clauses used as indirect speech acts needs to be accounted

3. Glozz is an annotation tool designed by Yann Mathet and Antoine Widlöcher. The annotation scheme relies on units, relations and schemas. <<http://www.glozz.org>>

4. Nonetheless, on the basis of experimental studies comparing enacted and naturally-induced vocal data, Scherer et al. (2011: 409) stress that the two procedures yield similar results, which suggests that "it may not matter very much whether emotional expressions are enacted or experimentally induced, at least for some major emotions".

5. Clarification requests straddle the border between direct and indirect speech acts. As shown by Celle et al. (forthc.), clarification requests may be used as indirect speech acts in a purely expressive way.

6. Embedded interrogatives are left aside in this paper. For a comparison of the uses of root interrogatives and embedded interrogatives, see Celle (2009).

for; so does the nature of the speaker-addressee relationship when no answer is requested. I first examine unresolvable questions before moving on to rhetorical questions.

3. Unresolvable questions

Unresolvable questions are the least frequent category of surprise-induced questions in our sample ($n = 5$). This category is based on the conjectural question type put forward by Littell et al. (2010) to account for the wonder effect produced by the insertion of a conjectural / inferential evidential into a question in three Amerindian languages. Littell et al. maintain that conjectural questions are wonder-like statements, although formally, they are *wh*-interrogatives. The claim made in this paper is that a similar wonder effect is produced by the insertion of emotive modifiers into a question in English. However, strictly speaking, these questions do not express a conjecture in the sense that they cannot be rephrased using “I surmise / I presume” + content clause.⁷ It is the proposition as a whole that is a matter of conjecture. In and of itself, this question type reflects the speaker’s ignorance rather than their conjecture. Therefore, I propose to dub these questions unresolvable rather than conjectural. Emotive modifiers indicate that the situation is appraised as violating the speaker’s expectations to such an extent that the question-answer presupposition is cancelled.

In the surprise contexts that were examined for the present study, unresolvable questions may contain interjections (such as ‘shit’, ‘gosh’), emotive modifiers (such as ‘on earth’ or ‘the hell’)⁸ or deictics, that is, items that mark the speaker’s emotional

7. In French, the inferential conditional used in an interrogative clause produces a conjectural question, and not an unresolved question: *Or cet enfant venait d’être volé par un inconnu. Quel pouvait être cet inconnu? Serait-ce Jean Valjean? [The child had just been stolen by an unknown man. Who could that unknown man be? Could it be Jean Valjean?]* (Hugo, 1862 Frantext). As stated by Dendale (2010: 297; 302), the “interlocutive function” of the question is affected by the conditional. However, the reason for the weakening of the interrogative force is that the speaker believes the proposition to be true. The conjectural question may be considered a mitigated assertion that can be rephrased as “I surmise that p” (*Je suppose que c’est Jean Valjean*). As shown by Diller (1977: 3–4), the conditional conveys a presupposition of evidence that is superimposed on the question, which reduces its interrogative force. She argues that a conjectural question in the conditional asserts a presupposition. I claim that a conjectural question seeks the addressee’s commitment (Celle 2007). By contrast, expressives in unresolvable questions are triggered by defective evidence. They implicate that no value can instantiate the question variable, which precludes assertion. This can be paraphrased as “I don’t know if p; I don’t know where / what / how ...”.

8. The distinction between interjections and emotive modifiers is borrowed from Huddleston & Pullum (2002: 916).

involvement and context-boundedness (see Ameka 1992: 108). Some of these items are swearwords ('gosh', 'shit', 'the hell') used cathartically (see Pinker 2007), that is, they serve an intra-individual function by reducing the stress associated with the utterance situation once it has been appraised as discrepant. Both interjections and emotive modifiers take the utterance situation as the source of surprise. As noted by Vingerhoets, (2013: 290), "cathartic swearing is regarded as an **adaptation**,⁹ especially meant to communicate that the situation we are confronted with deeply affects us, as evidenced by the display of strong emotions".

Adjustment to direct evidence is a feature shared by questions with emotive modifiers in English and conjectural questions in the languages that have evidentials or inferential conditionals. It gives credence to the claim that unresolvable questions are a cross-linguistic phenomenon that can be extended to English. Indeed, emotive modifiers point to defective evidence about the potential answers to the question, so that the addressee cannot be expected to provide an answer.

The fact that these highly emotional questions are systematically content questions suggests that a correlation can be established between speaker perspective and *wh*-questions (as shown by Celle & al. (forthc.) in the case of interrogatives used as direct speech acts, and by San Roque et al. (2017) in the case of evidential questions). This correlation is all the more striking as the most frequent questions in standard communication contexts in English are polar questions (see Stivers 2010; Siemund 2017). This suggests that the more emotional a question is, the more open-ended the set of answers will be. Unresolvable questions used in surprise contexts are about a salient open proposition. The clash between the speaker's expectations and the incongruous character of the situation makes it impossible for the speaker to assign a value to the question variable and to expect the addressee to be able to do so.

In Littell et al's (2010) typology, conjectural questions differ from rhetorical questions in that the speaker does not know the answer; they also differ from ordinary questions in so far as they do not require an answer from an addressee. This holds true for unresolvable questions in English:

- (1) **Ed:** Whoa, look at this camel, this is a real camel, **Gosh, where'd they get a real camel?**
- (2) **Bela:** Oh, there's my bus. [he checks his pockets] **Shit, where's my transfer?!**
Ed: Don't you have a car?
- (3) **EXECUTIVE 1** **What the hell is this?!**
EXECUTIVE 2 **Is this an actual movie?!**

9. My emphasis. The adaptation marked by swearing may be regarded as the verbal expression of the cognitive and emotional adaptation to unexpected events that underlies surprise (see Darwin 1872/1965).

These questions can be rephrased as follows:

| | |
|-----------------------|------------------------------|
| I don't know / wonder | where they got a real camel. |
| | where my transfer is. |
| | what this is. |

The source of wondering is the unexpected presence, absence or location of some element in the utterance situation. In other words, the cause of surprise is an extralinguistic event that violates the speaker's expectations. It is some new environmental information (see Peterson 2017) that may be surprising to both speaker and addressee. As such, it constitutes defective evidence, to the point that the experiencer cannot make inferences about the situation (Stein & Hernandez 2007: 302). The experiencer is forced to revise his / her previously held beliefs. In (1), there is a real camel on stage although the speaker assumes there should not be one; in (2), the transfer is not in the speaker's pocket although it should be there; in (3), the properties of the movie defy the speaker's ability to characterize it. At the same time, the specific contribution of the interjection or the emotive modifier is that evidence is so defective that the addressee is not expected to know the answer.

These questions may even be self-addressed as in (1), where there is no addressee. In (2) and (3), no answer is provided by the addressee who responds by asking a biased question or a rhetorical question, and the interchange is perfectly felicitous. In (3), the follow up polar question restricts the set of possible values for the question variable and specifies the nature and quality of the entity that both speaker and addressee find surprising.

Pragmatically, unresolvable questions are speaker-oriented, like exclamative utterances. Huddleston and Pullum (2002: 916) stress that emotive modifiers "express surprise or bafflement, and hence suggest that the speaker does not know the answer to the question. They tend to emphasise the open-endedness of the set of possible values for the question variable." Noteworthy is the fact that they may be followed by a question mark as well as an exclamation mark. However, these questions have the syntax and the semantics of interrogatives.

Syntactically, they require subject auxiliary inversion. Semantically, they are concerned with the identification and the appraisal of an incongruous situation,¹⁰ not with degree. Unlike most exclamatives, unresolvable questions carry no explicit "scalar implicature" (Michaelis & Lambrecht 1996: 378).¹¹ However, they do imply

10. The concept of "incongruity judgement" was originally coined by Kay and Fillmore (1999).

11. A parallel may be drawn here between unresolvable questions and the WXDY construction as defined by Kay and Fillmore (1999: 25–26). However, in the case of unresolvable questions, mirative meaning is related to their deictic nature. As shown by Kay and Fillmore, WXDY

an implicit scale by suggesting that the actual state of affairs violates the speaker's expectations or norms in an extreme way.

Unresolvable questions containing interjections may be differentiated from those containing emotive modifiers for two reasons. First, the valence associated with interjections may be either positive (1) or negative (2), while emotive modifiers tend to be associated with a negative valence (3). Second, emotive modifiers convey a stronger expressive meaning than interjections, which has implications on the function of the speech act. Unresolvable questions containing interjections allow continuation with an informative answer, although they do not request an answer:

- (1') A – Whoa, look at this camel, this is a real camel, Gosh, where'd they get a real camel?
 B – In the Sahara.
- (2') A – Oh, there's my bus. Shit, where's my transfer?!
 B – You must have left it at home.

As stated by Ameka (1992: 107) interjections “encode speaker attitudes and communicative intentions and are context-bound”. They express the speaker's emotional reaction to some unexpected event: the presence of a real camel in (1), the absence of the transfer in (2). The transfer is not where it is expected to be, the real camel is unexpected in this setting, hence the unresolvable questions about the origin of the camel in (1) and about the location of the transfer in (2). The answer may well increase the speaker's knowledge by assigning a value to the place variable in (1') and (2'). However, the aim of the question is not to increase the speaker's knowledge, but to express the emotional reaction of the speaker faced with an unexpected discrepant situation. Whatever the answer, it does not eliminate the sense of surprise.

Emotive modifiers provide questions with a strong expressive force, which overrides referential meaning. ‘The hell’ systematically follows the wh-word and the sequence “wh-word the hell” is a semi-fixed phrase. In (3), the speaker's question is triggered by visually perceived incongruous evidence. As the speaker is witnessing the situation, the addressee's answer is redundant in the sense that it does not increase the speaker's knowledge:

- (3') ‘What the hell is this?!’
 ‘This is an actual movie.’

constructions can express a sense of incongruity independently of the situation of utterance as they can be embedded. See also Celle & Lansari (2015) on *aller* + infinitive in one of its uses. This calls for further research into the connection between mirative constructions (possibly unrelated to the speaker's here and now) and mirative utterances (deictically related to the speaker).

Paradoxically, this referential answer is also insufficient because it fails to account for the incongruous character of the state of affairs. Providing a value for the variable is not enough to account for the incongruity of the state of affairs. The unresolvable question is not about the identification of the situation the speaker is witnessing. It conveys a negative assessment of the film movie that is being watched because the movie does not meet the standards of an actual movie. The answer can eliminate neither the sense of incongruity nor the negative assessment.

Huddleston and Pullum (2002: 916) rightly note that ‘which’ cannot substitute for ‘what’ when a what-question contains an emotive modifier. According to them, the open-endedness of the set of possible values implied by emotive modifiers accounts for that restriction: “as a result, they are hardly compatible with *which*, for this involves selection from an identifiable set”. This reasoning can be carried one step further. I argue that these questions are not concerned with the identification of a referential entity. The emotive modifier *the hell* cancels the question-answer presupposition by suggesting that whatever the answer, the situation violates the speaker’s expectations to such an extent that their state of surprise and negative assessment cannot be altered.¹² Given the extreme character of the situation, any informative answer is epistemically pointless. This type of question constitutes an expressive speech act devoid of any epistemic goals (see Zaefferer 2001: 224).

Interjections and emotive modifiers impart a mirative meaning to unresolvable questions, although mirativity is not encoded morphosyntactically in English. As stressed by DeLancey (2001: 377–378), mirativity is a “covert semantic category” in English, as opposed to other languages. However, interjections and emotive modifiers do encode the speaker’s surprise and relate it to new environmental information. Strikingly, all the unresolvable questions in our sample are induced by new environmental information and not by a surprising discourse entity. By contrast, questions used as direct speech acts in surprise contexts (i.e. clarification requests, ordinary questions and inferential questions) are mainly induced by a surprising discourse entity (see Celle & al. *forthc.*) and therefore do not qualify as mirative utterances (see Peterson 2017: 68).

12. In the surprise contexts studied in this paper, the *wh*-word-‘the-hell’ phrase indicates that the cause of surprise is the utterance situation. In such contexts, I argue that the phrase cancels the question-answer presupposition. However, the *wh*-word-*the-hell* phrase may be used in requests for information that simultaneously carry an instruction of unresolvedness especially when reference is made to a past event. In the following example borrowed from den Dikken and Giannakidou (2002: 32), the interrogative is an information question about the identity of the buyer: ‘Who the hell bought that book?’ As stated by den Dikken and Giannakidou (2002: 32), the *wh*-word-‘the-hell’ phrase conveys the presupposition that ‘Nobody was supposed to buy the book’. This results from the instruction of unresolvedness (i.e. the speaker’s failed attempt to resolve the question), which is nonetheless compatible with a genuine information question about the identity of the buyer.

The incompatibility of emotive modifiers (such as ‘the hell’, ‘on earth’) with echo questions Fillmore (1985: 82) and with which-questions observed by Huddleston and Pullum (2002: 916) and Pesetsky (1987: 111) substantiates this claim by suggesting that the surprising element cannot be traced back to the previous discourse. To quote Pesetsky (1987: 111), ‘the hell forces “a non-discourse-linked reading”, while “which-phrases are discourse-linked”. It is the context-boundedness of emotive modifiers that allows for the indexical mirative meaning of unresolvable questions.

The mirative meaning of this type of question can be probed using the undeniability test (Rett & Murray 2013: 455; Celle & al. 2017: 220):

- (3") A. ‘What the hell is this?!’
 B. #‘You are not surprised’.

The sense of surprise cannot be denied, which reveals that this utterance is an expressive speech act (see Potts 2005: 157). As such, the mirative speech act can only reflect the speaker’s emotional state, and the addressee cannot deny that emotional state.

Interestingly, Alcázar (2017: 37) points out that in Basque, the mirative particle *ote* often collocates with swearwords equivalent to ‘the hell’ in questions of the type ‘Can’t-find-the-value-of-x’. There seems to be typological evidence that the use of emotive modifiers in unresolvable questions pertains to mirativity. Like evidentials in some languages, emotive modifiers may shift the interpretation of questions, which take on an ignorance meaning.¹³ This shows that they have not only an intensifying or emphatic function in questions (Hoeksema & Napoli 2008). They also have an illocutionary effect on interrogatives.

4. Surprise-induced rhetorical questions

4.1 Expectation violation

Formally, rhetorical questions resemble questions used as direct speech acts. Surprise-induced rhetorical questions can enter into the same syntactic patterns as direct questions, except the declarative pattern. Rhetorical questions have an interrogative syntax in a more systematic way than ordinary questions. *Wh*-questions with subject-auxiliary inversion account for half of all the interrogative clauses used as indirect speech acts in my sample. The different types of rhetorical questions are classified in Table 1.

13. However, more research is needed to elucidate the relation between indexicals, evidentials and emotive modifiers. Some scholars (see for example Korotkova 2016: 224–226) argue that evidentials are addressee-oriented in questions, in contrast to indexicals which remain speaker-oriented.

Table 1. Types of rhetorical questions

| | | |
|--------------------------------------------------------------------|-----------------------------------------------------------------------|---|
| <i>Wh</i> -question: <i>Wh</i> -word + subject-auxiliary inversion | What kind of sick mind would operate like that? | 5 |
| Yes/no question | Can you imagine what that guy would be like in a movie? | 4 |
| Reprise fragment ^a | Stronger? You see! You see!! You stupid minds! Stupid! | 1 |
| <i>Wh</i> - question: clause fragment | And what about this so-called “Barbara” character? It’s obviously ME! | 1 |
| Reprise sluice | Since when? | 1 |

^a This term is borrowed from Ginzburg (2012).

In total, *wh*-questions (including a reprise sluice and a verbless sentence) are found in 8 out of 12 rhetorical questions. This confirms the correlation established above as well as in Celle et al. (forthc.) between surprise and *wh*-questions, that is, questions that denote a set of possible answers. *Wh*-words include ‘what’ ($n = 3$), ‘when’ ($n = 2$), ‘how’ ($n = 3$). ‘How’-questions are always associated with a modal auxiliary in their rhetorical reading. Among the elements that facilitate a rhetorical reading are also deictic items (‘like that’) and degree words (‘so casual’).

Semantically, rhetorical questions define either a closed set of possible answers or an open set of possible answers, the latter being the most frequent case in my sample. Informationally, rhetorical questions are complete utterances, as opposed to ordinary questions. Pragmatically, they express a biased position by pointing towards an obvious answer (Rohde 2006: 149). As pointed out by Caponigro and Sprouse (2007: 131), “[r]hetorical questions are not asked to trigger an increase in the amount of mutual knowledge”, nor do they assert anything new. This raises two questions that are addressed below. First, how can rhetorical questions qualify as questions? Second, why are rhetorical questions used to express surprise? Such expressions reveal an epistemic asymmetry, the speaker’s expectations conflicting with the addressee’s beliefs or with the state of affairs. In her analysis of responses to rhetorical questions, Rohde (2006: 142) notes that rhetorical questions “generate very little surprise”: “[T]he case of complete lack of surprise corresponds to rhetorical questions because the answer is predictable to both the Speaker and the Addressee. The answer is so unsurprising that it need not be uttered at all.” (ibid: 147).

My analysis of responses to rhetorical questions yields similar results: the answer to the rhetorical questions found in surprise contexts need not be uttered.¹⁴

14. Rohde (2006) uses naturally-occurring language data drawn from the Switchboard corpus, a corpus of telephone speech. Her approach is purely semasiological – i.e. her analysis is not focused on rhetorical questions occurring in surprise contexts.

However, it may not be because the answer is unsurprising. Under Rohde's analysis (2006: 152), the fact that rhetorical questions generate little surprise is "evidence of their unformativity". Adopting Gunlogson's (2001) Common Ground theoretical framework, Rohde (2006: 152) views this unformativity as indicative of the fact that rhetorical questions "require no update to participants' commitment sets". This view is challenged in the present chapter. It is also argued that informativity and update of the participants' commitment sets should be distinguished. In surprise-generated rhetorical questions, no informative answer is requested, but rhetorical questions are uttered in reaction to some unexpected linguistic information or incongruous situation and involve a two-fold update. First, they signal the speaker's attempt to cognitively integrate unexpected new information. Second, they request a commitment update on the part of the addressee. With respect to surprise, Rohde only examines responses to rhetorical questions without considering how these questions may lend themselves to the expression of surprise. I contend that rhetorical questions may be used to express surprise precisely because the nature of the speech act they convey allows for the expression of conflicting views in a questioning process whereby the addressee is asked to update his / her commitment.

My claim is that this pragmatic commitment update process is highly congruent with the appraisal process that underlies the surprise reaction on the psychological level. Within a schema-theoretic framework, Meyer et al. (1997: 253) characterize the surprise-induced appraisal process as follows:

[S]urprise-eliciting events initiate a series of processes that **begin** with the appraisal of a cognized event as exceeding some threshold value of schema-discrepancy (or unexpectedness), **continue** with the occurrence of a surprise experience and, simultaneously, the interruption of ongoing information processing and reallocation of processing resources to (i.e. the focusing of attention on) the schema-discrepant event, and **culminate** in an analysis and evaluation of this event plus – if deemed necessary – an updating, extension, or revision of the relevant schema.

The questioning process encoded by rhetorical questions necessarily requests a commitment update from the addressee. Furthermore, surprise-induced rhetorical questions appear to be much more complex emotionally than surprise-induced ordinary questions. Surprise may be tinged with anger in rhetorical questions that are typically asked to express disbelief and disagreement. In that case, expectation violation is coupled with the violation of standards and the thwarting of the experiencer's goals, which correspond to the ingredients of anger as defined by Ortony and al. (1988: 152–153).

4.2 Informative answers

Rhetorical questions are generally said to be semantically equivalent to statements because they contain the answer to the question they ask and do not request an answer from the addressee. The view upheld in the present chapter is that rhetorical questions necessitate a pragmatic account. Even when the rhetorical intent is obvious from the pragmatic context and the semantic construction, the addressee may fail or deliberately refuse to recognize it. In that case, an informative answer may be provided:

- (4) **Reverend Lemon:** Mr. Wood? What do you think you're doing?!
Ed: I'm directing.
Reynolds: Not like THAT, you're not!
- (5) **Rachel:** I'm allergic to peanut butter.
Ray: (laughs) Since when?
Rachel: (with a snotty look) Birth!

These questions fail the tests designed by Caponigro & Sprouse (2007) to reveal information-seeking questions:¹⁵

- (4') **#I'm really curious:** What do you think you're doing?!
#I really don't know: What do you think you're doing?!
- (5') **#I'm really curious:** Since when (have you been allergic to peanut butter)?
#I really don't know: Since when (have you been allergic to peanut butter)?

In addition to these tests, rhetorical meaning is also revealed by certain interrogative phrases. *When* as a complement of the preposition *since* is less likely to request an informative answer that selects a temporal starting point. It suggests a sudden start that may not be relevant to some states, such as being allergic.¹⁶ As noted by Huddleston & Pullum (2002: 905), “‘since when’ is often used sarcastically, with cancellation of the presupposition”. In the following example, ‘since when’ points to a rhetorical reading in a similar way:

- (6) **Robbie:** I don't have a license.
Ray: Since when has that stopped you?

The ‘since when’-question is not about the starting point of the stopping process but cancels the question-answer presupposition ‘That has stopped you for some

15. Except in sarcastic contexts where they are felicitous, as pointed out by an anonymous reviewer.

16. As rightly pointed out by an anonymous reviewer.

time'. The rhetorical question implies 'That has never stopped you'. Consequently, Robbie's statement 'I don't have a license' loses its argumentative force as a justification for not driving.

Contrast with a 'how long'-question:

- (7) A – I'm allergic to peanut butter.
 B – How long have you been allergic to peanut butter?

This question is accompanied by a question-answer presupposition ('you have been allergic to peanut butter for some time') and seeks to assign a value to a variable in an open proposition ('you have been allergic to peanut butter since X').

Although the questions in (4) and (5) are unequivocally rhetorical, they are followed by an informative answer. Such answers are not requested. However, they are perfectly compatible with rhetorical questions syntactically and conversationally because rhetorical questions are questions, not assertive statements. They are made possible by the semantic nature of these questions, which denote a set of potential answers ('playing', 'working', 'directing', etc. in (4), 'since birth', 'since 1960', 'since the war', etc. in (5)). Nonetheless, such answers refute the rhetorical scenario on the pragmatic level and express disagreement. Instead of acknowledging the rhetorical intent whereby the speaker points to some answer supposedly obvious to both speaker and addressee (i.e. 'whatever you think you're doing is wrong, you have never been allergic to peanut butter'), the addressee may well choose to assign a value to the variable as if the question were information-seeking.

According to Rohde (2006: 161), rhetorical questions are understood as such by virtue of their properties of "answer obviousness" and "similarity", that is, the answer to a rhetorical question is obvious to speaker and addressee and they supposedly both share a commitment to an answer similar in nature. However, an unco-operative addressee may reject or ignore the rhetorical effect imposed by the speaker even if it is recognized as such. It is noteworthy that informative answers are found in the case of second-person utterances – or at least in questions involving the second person as in the reprise sluice in (5). These rhetorical questions are indirect speech acts that challenge either what the addressee is doing (as in (4)), or what the addressee has just said (as in (5)) because speaker and addressee do not share the same standards or the same beliefs. The sense of absurdity conveyed by the rhetorical questions may not be shared by the addressee. Providing an informative answer amounts to assigning a value to a variable as in the case of an ordinary question. In this way, the addressee avoids committing to the proposition that is indirectly asserted by the rhetorical question ('You are messing up' in (4), 'You have never been allergic to peanut butter' in (5)). The rhetorical question then fails to update the addressee's commitment.

4.3 The surprise-induced rhetorical scenario

Rhetorical questions may be followed by an informative answer, although they do not invite such an answer. More often than not, they are followed by a response. Each case is examined in turn.

- (8) **Dolores:** Ugh! How can you act so casual, when you're dressed like that?!
Ed: It makes me comfortable.

The answer can be construed as the causal explanation for acting so casually. It indicates that the question is taken to carry a presupposition ('you act very casually when you're dressed like that') while the implied assertive statement, that is, the implicit evaluative judgment implied by the question ('you shouldn't act so casual when you're dressed like that'), is ignored, which foils the rhetorical strategy. The rhetorical strategy fails in a similar way in the following constructed examples:

- (9) **A** – Ugh! How can you act so casual, when you're dressed like that?!
B – Thanks to my talent.
- (10) **A** – How can you just walk around like that, in front of all these people?
B – With a walking stick.
- (11) **A** – Goldie, how many times have I told you guys that I don't want no horsin' around on the airplane?
B – Just once.

These answers signal that the addressee deliberately ignores the rhetorical scenario imposed by the speaker. By contrast, responses do not attempt to undermine the rhetorical strategy, but take disagreement for granted:

- (12) **Dolores:** How can you just walk around like that, in front of all these people?
Ed: Hon', nobody's bothered but you.
- (13) **Kong:** Goldie, how many times have I told you guys that I don't want no horsin' around on the airplane?
Goldie: I'm not horsin' around, sir, that's how it decodes.

In (12) and (13), the responses signal the addressee's disagreement with the evaluative judgments expressed in the rhetorical questions. As such, these rhetorical questions force the addressee to accept the implied assertive statement, and do not request a response. First, I define the nature of the discrepancy conveyed by rhetorical questions before taking up the issue of addressee commitment.

Surprise-induced rhetorical questions express a conflict between the speaker's epistemic domain and the actual state of affairs: in (13), the rhetorical question implies an indirect assertive statement: 'I have told you so many times that I don't want no horsin' around on the airplane'. The source of the speaker's surprise is the

soldier's preceding answer, which confirms surprising information. This answer is mistakenly construed as a joke, i.e. as an act of disobedience. The rhetorical question indicates that the speaker is epistemically unprepared to face a totally unexpected turn of events and is therefore unable to behaviorally adapt to it. Discrepancy arises from the speaker's failure to correctly interpret an unexpected answer. Some fact is directly perceived but misinterpreted.

When rhetorical questions contain modal auxiliaries, they typically express a conflict between realis and irrealis:

- (14) What kind of sick mind would operate like that?

The modal 'would' occurs with question-answer presupposition cancellation, that is, the presupposition that some value can be supplied for the subject variable is cancelled by modality.¹⁷ The rhetorical question implies an indirect assertive statement, namely that 'no sound mind would operate like that'. In reaction to an actual situation for which the speaker is epistemically and morally unprepared, a hypothetical stance is adopted that strips the surprising situation of its realis quality. Surprise is related to the irrealis domain (see Akatsuka 1985). This sense of reality denial is particularly striking with *how*-rhetorical questions, which always contain the modal auxiliary *can* in my sample (as in (8) or (12)).

As noted by Desmets and Gautier (2009: 109), *comment* 'how'-rhetorical questions in French (such as *Comment peux-tu déambuler de cette façon?*) contain two contradictory pieces of information (*tu déambules de cette façon* and *tu ne peux pas déambuler de cette façon*). The same analysis holds for *how*-rhetorical questions in English. Indeed, (8) and (12) carry a presupposition that p ('you are walking around like that, with such accessories' in (12), and 'you are acting so casual' in (8)) that conflicts with the negative comment implied by the rhetorical questions ('you can't walk around like that, in front all these people' in (12) and 'you can't act so casual, when you're dressed like that' in (8)).

Counterfactual evidence runs counter to the assertion of p. It is supplied by a variety of markers, such as a temporal *when*-clause, a degree modifier ('so casual'), deictic items ('like that'), and a spatial PP ('in front of' ...). These markers all signal a violation of the speaker's expectations and are conducive to the indirect assertion of non (modality) p, although p is presupposed.¹⁸ This discrepancy accounts for the mirative meaning of these rhetorical questions.

17. For a detailed account of *would* in questions, see Celle (in press) and Celle & Lansari (2014) and (2016).

18. This notation is borrowed from Desmets and Gautier (2009). They argue that *comment*-rhetorical questions in French conflate a question about the modal operator (*pouvoir*) of the proposition and an assertion that negates both the modal operator and the proposition. *How*-rhetorical questions in English with the modal auxiliary 'can' behave in a similar way.

4.4 Addressee commitment

In surprise contexts, rhetorical questions may be considered mirative utterances¹⁹ not only because they express surprise, but also because the discrepancy they convey triggers a specific stance on the part of the speaker. Modalised rhetorical questions deny reality either by cancelling the question-answer presupposition²⁰ or by relying on counterfactual evidence.²¹ The speaker directly perceives some event, which should lead to the assertion of *p*, the status of *p* being in no doubt. And yet, the speaker does not commit to the truth of *p*, because *p* runs counter to her expectations. The function of rhetorical questions is then to question the grounds that made *p* possible. In a rhetorical question, the speaker selects an answer and requests the addressee to commit to the truth of that proposition.²² Rhetorical questions are biased because they do not leave any choice to the addressee with respect to the selection of a variable. However, they are questions in the sense that they request the addressee's commitment to a proposition. They can be preceded by the discourse marker *tell me*, which, as shown by Reese (2007: 51) co-occurs with questions that request a response, but not with assertions. Asked to commit to a proposition that stands in contrast to the state of affairs or to her beliefs, the addressee may not respond:

- (15) **Ed:** This is my way of telling you –
Dolores: [furious] What, by putting it in a fuckin' script, for everyone to see?!
What kind of sick mind would operate like that?
 [Ed is terribly hurt. Dolores shakes that script.]

19. Mirativity is defined by several authors (among others, Guentchéva (2017), Peterson (2017)) as resulting from a discrepancy between what is observed and what is expected.

20. On cancellation of question-answer presupposition, see Huddleston and Pullum (2002: 900–901).

21. My translation of “indice contrefactuel”, a concept borrowed from Desmets and Gautier (2009: 111).

22. I agree with Desmets & Gautier (2009) and Beyssade & Marandin (2009) that the speaker requests the addressee's commitment to a proposition in rhetorical questions. However, rhetorical questions do not necessarily request a verbal response, let alone an answer. They seek the addressee's alignment, but it is not clear whether the addressee will actually commit to the proposition suggested by the speaker. If rhetorical questions are not generated by surprise, anger or disagreement, the assertion they imply is uncontroversial and the answer to the question is obvious. The absence of an answer from the addressee may then be interpreted as tacit agreement. In emotion-induced rhetorical questions, however, a commitment update on the part of the addressee is requested. In half of the examples of my sample, rhetorical questions are not followed by a response or an answer. In the absence of any explicit verbal reaction from the addressee, it is difficult to determine whether the speaker and addressee's common ground is eventually updated.

Dolores: And what about this so-called 'Barbara' character? It's obviously ME! I'm so embarrassed! This is our life!

The rhetorical question allows the speaker to express disapproval without committing to the proposition 'You are a sick mind'. The addressee is asked to commit to the proposition 'No sound mind would operate like that'. This indirect insult may reach its goal emotionally, as specified in the stage direction ('Ed is terribly hurt'). From an interactional perspective, however, it is a dead end.²³ Unless an unlikely response such as 'You are right' or 'I know' is uttered, no commitment update is possible on the part of the addressee.

In (8) and (12) a response is provided by the addressee, but it does not update the common ground in the way expected by the speaker: what is presented as counterfactual evidence according to the speaker's standards is said to be normal behavior in the addressee's response. As a result, the contradiction conveyed by the rhetorical question is cancelled and the speaker is forced to accept *p*, even if the rhetorical question requests the addressee to commit to non (modality) *p*. In (13), the addressee's response forces the speaker to revise his appraisal of the situation.

The analysis of emotion-induced rhetorical questions shows that the response to a rhetorical question is not obvious to both speaker and addressee when they have different expectations. Contra Rohde (2006: 149–150), I argue that the addressee may be committed to a proposition that contradicts the speaker's bias. The common ground may then be updated, but in a way that is not anticipated by the speaker, especially when the speaker denies an unexpected but actual event that *s/he* has failed to cognitively integrate.

The mirative nature of emotion-induced rhetorical questions has important theoretical implications. As argued by Alcázar (2017: 37), mirative rhetorical questions express "antithesis of the Common Ground", which goes against standard treatments of rhetorical questions. In line with Alcázar, I believe that the deictic essence of the rhetorical questions under study accounts for their mirative meaning.²⁴ Evidence of this deictic component may be supplied by their resistance to

23. Although surprise-induced rhetorical questions may take different forms and do not as such constitute a construction, a parallel may be drawn here with the Split Interrogative construction analysed by Michaelis and Feng (2015). The conversational dead end is typical of what Michaelis & Feng (2015: 149) call sarcastic syntax. Under their analysis, ironic utterances are "counterfeit speech acts" that "do not advance the conversation" because their function is "disruptive".

24. I view mirativity as an epistemic stance adopted in reaction to an unexpected event. Mirativity consists in the expression of surprise, not in the description or assertion of surprise (Celle & al. 2017). Along this line of reasoning, expressions like 'I am surprised, 'he was surprised' are not mirative utterances. These are assertions of surprise (see Rett & Murray 2013: 455). They need not be anchored to the time of utterance or to the first and second persons. Such expressions can be embedded without any change in the surprise meaning.

the embeddability test. Mirative meaning is lost in an embedded clause (see Rett and Murray 2013, Celle et al. 2017):

(15') Dolores asked Ed what kind of sick mind would operate like that.

This embedded sentence cannot express Dolores's surprise, contrary to the rhetorical question in (15). Even if this sentence is turned into an exclamation, it can only express the speaker's surprise, not Dolores's:

(15'') Dolores asked Ed what kind of sick mind would operate like that!

However, it should be stressed again that in English, mirativity is not marked as such morphosyntactically.²⁵ Predictably, the form and structure of mirative rhetorical questions do not differ from those of non-mirative rhetorical questions. Combining a semasiological approach with an onomasiological perspective offers a means to detect a meaning that might otherwise go unnoticed for lack of a dedicated morpheme. It also enables us to enrich our understanding of mirativity in English. DeLancey (2001: 377–378) suggests that mirativity is a “covert semantic category” in English mainly expressed intonationally.²⁶ I further argue that rhetorical questions, i.e. questions used as indirect speech acts, may serve a mirative function. One might wonder why there is such an affinity between mirativity and rhetorical questions. From a schema-theoretic perspective, surprise induces cognitive processing that starts with the search for a cause and ultimately ends with belief revision (see Meyer et al. 1997: 253; Miceli & Castelfranchi 2015: 52). This highly adaptive psychological pattern is ideally mapped on rhetorical questions, which call upon the addressee for commitment update to validate belief revision. Note that mirativity and rhetorical questions have non-commitment in common. The epistemic stance adopted by a speaker in reaction to unexpected information is typically one of non-commitment, as shown by Zeisler (2017) on the use of *ħdug* in Ladakhi.²⁷ Zeisler stresses that “SPEAKER ATTITUDE (OR STANCE) primarily

25. Nonetheless, there is typological evidence in support of mirative rhetorical questions. In Basque, for example, mirative rhetorical questions are marked by *ote*, a mirative conjunction (see Alcázar 2017). In Ashéninka Perené, Mihás (2014: 213–216) also shows that the enclitic *=ma~=taima* that commonly encodes inference can occur in content questions to express mirative meaning, both in direct and indirect speech acts (including rhetorical questions).

26. Mirativity in English is often associated with an exclamation intonation (Rett & Murray 2013). Questions used as indirect speech acts do have an exclamation intonation and can convey mirative meaning. As explained in Celle et al. (forthc.), surprise in questions used as direct speech acts is generally induced by a discourse entity rather than by new environmental information. Although such questions do express surprise, they should not be considered mirative.

27. *ħdug* is an auxiliary used to encode visual perception and non-commitment. Zeisler (2017) argues that *ħdug* has parasitic mirative connotations.

deals with the relation between the speaker and the content of the utterance *and* between the speaker and the addressee". Rhetorical questions are also primarily addressee-oriented as they request the addressee's commitment, the speaker putting forward an answer without committing to the truth of the proposition.

This leads us to refine the epistemically-based typology of questions proposed in Littell et al. (2010). Questions are not only based on the speaker's knowledge and beliefs.²⁸ They also contribute to dialogue in a dynamic way by requesting the addressee's commitment. Therefore, an epistemically-based typology of questions should accommodate the request for commitment update that distinguishes rhetorical questions from assertions. In a rhetorical question, the speaker believes that the addressee knows the answer because the answer is suggested by the question itself, although speaker and addressee may differ in their beliefs and appraisals.²⁹ In the case of mirative rhetorical questions, the commitment update cannot be taken for granted as it may be hindered by disagreement. These features are summed up below in Table 2.

Table 2. An epistemically- and dialogically-based typology of questions

| | Speaker knows the answer | Speaker believes that the addressee knows the answer | Speaker requests addressee's commitment |
|------------------------|--------------------------|------------------------------------------------------|-----------------------------------------|
| Ordinary Questions | No | Yes | Yes |
| Rhetorical Questions | Yes | Yes | Yes |
| Unresolvable Questions | No | No | No |

5. Conclusion

This chapter is part of a study of all question types used in reaction to surprising information. As shown in Celle & al. (forthc.) 83% of the interrogatives found in our corpus of movie scripts are questions used as direct speech acts. These questions request information that aims to increase the speaker's knowledge and are generally discourse-linked, that is, surprise is generated by unexpected new linguistic information.

This chapter focused on the remaining 17% surprise-induced questions used as indirect speech acts. It showed that in English, mirativity can be conveyed by two different types of interrogatives used as indirect speech acts. Both surprise-induced

28. In addition, speaker's knowledge is grounded in dialogue and articulated to indexical cues, as shown by Du Bois (2007: 157).

29. By contrast, in an ordinary question, the addressee is asked to commit to her own answer.

rhetorical questions and unresolvable questions take the form of interrogatives but do not request information. Rhetorical questions are resolved questions, whereas unresolvable questions implicate that no resolution can be reached. Their mirative meaning is conveyed by a discrepancy between what is observed and what was expected. However, being generated by different types of surprising situations, these questions exhibit mirative meaning at different stages of the cognitive assimilation of unexpected new information.

Surprise-induced rhetorical questions are typically generated by counterfactual evidence. By using such an indirect speech act, the speaker distances him/herself from the discrepant actual state of affairs. Rhetorical content questions are found in reaction to some behavior for which the addressee is held responsible. Their aim is to persuade the addressee to modify that behavior in order to meet the speaker's expectations. From a schema-theoretic perspective, rhetorical questions may be viewed as the last stage of the surprise-induced appraisal process (see Meyer et al. 1997: 253 cited above): they aim at a commitment update on the part of the addressee. Using concepts borrowed from the schema-theoretic framework not only allows bridging the gap between psychology and linguistics. It also sheds light on the shades of mirativity by correlating them with different stages of the cognitive processing of new information induced by surprise. Rhetorical questions appear to express mirativity at the semantic-pragmatic level as mirative meaning is not associated with a specific morphosyntactic form in English. However, a constellation of grammatical and lexical items (modal auxiliaries, deictics, degree words) is the hallmark of mirativity.

Surprise-induced unresolvable questions tend to be triggered by directly perceived evidence. The addressee's agency is not involved – at least in my data – and a judgment of incongruity is formed once the situation has been appraised either positively or negatively. Whatever the answer – if any – it cannot account for the sense of incongruity attached to the situation. This type of expressive speech act does not carry an epistemic goal. Rather, it carries an instruction that no variable can be provided to instantiate the salient open proposition. Mirativity projects from the initial stage of the cognitive assimilation of unexpectedness: “the appraisal of a cognized event as exceeding some threshold value of schema-discrepancy (or unexpectedness)” (Meyer et al. 1997: 253). At that stage, the surprise process produces a state of ignorance and wonder. It is expressed by specific lexemes, that is, interjections and emotive modifiers that provide the indirect speech act with an expressive force.

Pragmatically, rhetorical questions express a biased position and are generally said to point to an obvious answer. The present chapter offers an alternative analysis that accounts for the apparent paradox of rhetorical questions being used in reactions of surprise. Contra Rohde (2006), I argue that informativity and update

of the participants' commitment sets should be distinguished. In surprise-generated rhetorical questions, although no informative answer is requested, a two-fold update is expected. First, rhetorical questions signal the speaker's attempt to cognitively assimilate new environmental information, the actual state of affairs being counterexpectational. Second, they request a commitment update on the part of the addressee in a questioning process triggered by the speaker's and the addressee's conflicting views. Rhetorical questions show that surprise, in association with other emotions it contributes to generating (such as anger), can be exploited within complex argumentative strategies, as evidenced in other research works on the lexicon of surprise (see Tutin 2017; Celle et al. 2017). Emotion-induced rhetorical questions serve an argumentative function whereby the addressee is asked to commit to a proposition that the speaker does not commit to in a direct way. Rhetorical questions offer a pragmatic means attempt to reduce the belief discrepancy associated with the experience of surprise. However, if the belief discrepancy cannot be reduced - in case of strong disagreement - they take on a challenging function.

Unlike rhetorical questions, unresolvable questions are generated by evidence judged incongruous. They are speaker-oriented, the speaker attempting to emotionally adapt to an incongruous situation without expecting an answer or even a response of the addressee. In English, interjections and emotive modifiers encode the mirative meaning of unresolvable questions. Further investigations are needed to better assess their respective contributions to mirativity. The claim made in this chapter is that in English, expressives can change the illocutionary force of a sentence in the same way as evidentials in other languages. This can be explained on the grounds that evidentials and expressives share common features. Emotive modifiers and interjections are illocutionary modifiers triggered by direct evidence. They encode the speaker's emotional experience, while evidentials encode the speaker's perceptual or cognitive experience. In sum, both evidentials and expressives encode speaker perspective.³⁰ Further investigations are needed to better assess the respective contributions of expressives and evidentials to interrogatives.

This chapter offers a refinement of Littell et al.'s (2010) typology of questions by including the commitment update parameter. It also proposes to distinguish between conjectural questions and unresolvable questions. In conjectural questions, the speaker knows the answer and only seeks the addressee's commitment to the truth of the proposition. In unresolvable questions, emotive modifiers change the illocutionary force by implicating that neither speaker nor addressee can provide an answer.

30. See San Roque et al. (2017). Potts (2007: 173) argues that expressives may be embedded and involve "perspective dependence" rather than strictly speaker perspective.

These findings also suggest that unresolvable questions are generated by outcome-related surprise, while rhetorical questions are generated by person-related surprise.³¹ I leave it to future research to determine whether this distinction is reflected in the appraisal pattern and whether it generates differences in linguistic responses. The complex relation between evidence-induced mirative utterances and discourse-based topic-comment constructions is also an avenue for future research.

Acknowledgements

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007–2013) under REA grant agreement n. PCOFUND-GA-2013-609102, through the PRESTIGE programme coordinated by Campus France. This paper was made possible by the data annotation carried out with Hakima Benali, Anne Jugnet, Laure Lansari and Emilie L'Hôte. I wish to thank Anne Jugnet, Laure Lansari and Tyler Peterson for the discussions we had on questions. I am especially grateful to Anne Jugnet and Laure Lansari for their feedback on previous drafts. I would also like to warmly thank the three anonymous reviewers for their valuable comments and suggestions. Any remaining errors are my own.

Corpus data

Base textuelle FRANTEXT, ATILF – CNRS & Université de Lorraine. Site internet : <http://www.frantext.fr>. Version décembre 2016.

<http://www.opensubtitles.org/>

Jörg Tiedemann. 2009. News from OPUS – A Collection of Multilingual Parallel Corpora with Tools and Interfaces. In N. Nicolov and K. Bontcheva and G. Angelova and R. Mitkov (eds.) *Recent Advances in Natural Language Processing* (vol V), pages 237–248, John Benjamins, Amsterdam/Philadelphia.

31. Van Dijk and Zeelenberg (2002) distinguish between outcome-related vs. person-related disappointment. Ortony et al. (1988) make a similar distinction between event-based and agent-based emotions. This distinction can be fruitfully applied to surprise and helps account for the connection between surprise and anger observed in the case of rhetorical questions, anger being a person-related emotion.

References

- Akatsuka, Noriko. 1985. Conditionals and the epistemic scale. *Language* 61(3): 625–639. doi:10.2307/414388
- Alcázar, Asier. 2017. A syntactic analysis of rhetorical questions. In *Proceedings of the 34th West Coast Conference on Formal Linguistics*, Aaron Kaplan, Abby Kaplan, Miranda K. McCarvel & Edward J. Rubin (eds), 32–41. Somerville MA: Cascadilla Proceedings Project.
- Ameka, Felix. 1992. Interjections: The universal yet neglected part of speech. *Journal of Pragmatics* 18(2): 101–118. doi:10.1016/0378-2166(92)90048-G
- Azzopardi, Sophie & Bres, Jacques. 2014. Futur, conditionnel, et effets de sens de conjecture et de rejet en interrogation partielle. *SHS Web of Conferences* 8: 3003–3013. doi:10.1051/shsconf/20140801014
- Beyssade, Claire & Marandin, Jean-Marie. 2009. Commitment: Une attitude dialogique. *Langue Française* 162: 89–107. doi:10.3917/lf.162.0089
- Bourova, Viara & Dendale, Patrick. 2013. Serait-ce un conditionnel de conjecture? Datation, évolution et mise en relation des deux conditionnels à valeur évidentielle. *Cahiers Chronos* 26: 183–200.
- Caponigro, Ivano & Sprouse, Jon. 2007. Rhetorical questions as questions. In *Proceedings of Sinn und Bedeutung 11*, E. Puig-Waldmüller (ed.), 121–133. Barcelona: Universitat Pompeu Fabra.
- Carroll, James M. & Russell, James A. 1997. Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology* 72: 164–176. doi:10.1037/0022-3514.72.1.164
- Celle, Agnès. 2007. Analyse unifiée du conditionnel de non prise en charge en français et comparaison avec l'anglais. *Cahiers Chronos, Etudes Sémantiques et Pragmatiques sur le Temps, l'Aspect et la Modalité* 19: 43–61.
- Celle, Agnès. 2009. Question, mise en question: la traduction de l'interrogation dans le discours théorique. *Revue Française de Linguistique Appliquée* XIV(1): 39–52.
- Celle, Agnès. 2018. Epistemic evaluation in factual contexts in English. In *Epistemic Modality and Evidentiality in a Cross-Linguistic Perspective*, Zlatka Guentchéva (ed.), 22–51. Berlin: Mouton de Gruyter.
- Celle, Agnès & Lansari, Laure. 2014. Uncertainty as a result of unexpectedness. *Language and Dialogue* 4(1) 7–23. doi:10.1075/ld.4.1.01cel
- Celle, Agnès & Lansari, Laure. 2015. On the mirative meaning of *aller* + infinitive compared with its equivalents in English. *Cahiers Chronos, Taming the TAME Systems* 27: 289–305. doi:10.1163/9789004292772_015
- Celle, Agnès, & Lansari, Laure. 2016. L'inattendu et le questionnement dans l'interaction verbale en anglais: Les questions en *why-would* et leurs réponses. In *Représentations du sens linguistique: Les interfaces de la complexité*, Olga Galatanu, Ana-Maria Cozma & Abdelhadi Bellachhab (eds), 235–248. Bern: Peter Lang.
- Celle, Agnès, Jugnet, Anne, Lansari, Laure & L'Hôte, Emilie. 2017. Expressing and describing surprise. In *Expressing and Describing Surprise* [Benjamins Current Topics 92], Agnès Celle & Laure Lansari (eds), 215–244. Amsterdam: John Benjamins.
- Celle, Agnès, Jugnet, Anne, Lansari, Laure & Peterson, Tyler. Forthcoming. Interrogatives in surprise contexts in English. In *Surprise at the Intersection of Phenomenology and Linguistics*, Natalie Depraz & Agnès Celle (eds). Amsterdam: John Benjamins.
- Darwin, Charles. 1965. *The Expression of the Emotions in Man and Animals*. Chicago IL: University of Chicago Press.

- DeLancey, Scott. 2001. The mirative and evidentiality. *Journal of Pragmatics* 33: 369–382.
doi:10.1016/S0378-2166(01)80001-1
- Dendale, Patrick. 2010. Il serait à Paris en ce moment. Serait-il à Paris? A propos de deux emplois épistémiques du conditionnel: Grammaire, syntaxe, sémantique. In *Liens linguistiques. Études sur la combinatoire et la hiérarchie des composants*, Castro Camino Alvarez, Flor María Bango de la Campa & María Luisa Donaire (eds), 291–317. Bern: Peter Lang.
- den Dikken, Marcel & Giannakidou, Anastasia. 2002. From hell to polarity: “Aggressively non-D-linked” wh-phrases as polarity items. *Linguistic Inquiry* 33(1): 31–61.
doi:10.1162/002438902317382170
- Desmets, Marianne & Gautier, Antoine. 2009. “Comment n’y ai-je pas songé plus tôt?” Questions rhétoriques en *comment*. *Travaux de Linguistique* 58: 107–125. doi:10.3917/tl.058.0107
- Diller, Anne-Marie. 1977. Le conditionnel, marqueur de dérivation illocutoire. *Semantikos* 2: 1–17.
- DuBois, John. 2007. The stance triangle. In *Stancetaking in Discourse* [Pragmatics & Beyond New Series 164], Robert Englebretson (ed.), 139–182. Amsterdam: John Benjamins.
doi:10.1075/pbns.164.07du
- Fillmore, Charles. 1985. Syntactic intrusions and the notion of grammatical construction. In *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*, 73–86. Berkeley CA: BLS.
- Ginzburg, Jonathan. 2012. *The Interactive Stance: Meaning for conversation*. Oxford: OUP.
doi:10.1093/acprof:oso/9780199697922.001.0001
- Guentchéva, Zlatka. 2017. An enunciative account of admirativity in Bulgarian. Special issue *The Linguistic Expression of Mirativity. Review of Cognitive Linguistics* 15(2): 539–574.
- Gunlogson, Christine A. 2001. *True to Form. Rising and Falling Declarative as Questions in English*. PhD Dissertation, University of California, Santa Barbara.
- Huddleston, Rodney & Pullum, Geoffrey. 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Haillet, Pierre-Patrick. 2001. A propos de l’interrogation totale directe au conditionnel. In *Le conditionnel en français*, Patrick Dendale & Liliane Tasmowski (eds.), 295–330. Paris: Klincksieck.
- Hoeksema, Jack & Napoli, Donna Jo. 2008. Just for the hell of it: A comparison of two taboo-term constructions. *Journal of Linguistics* 44: 347–378. doi:10.1017/S002222670800515X
- Jörg Tiedemann, 2009, News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces. In N. Nicolov and K. Bontcheva and G. Angelova and R. Mitkov (eds.) *Recent Advances in Natural Language Processing* (vol V), pages 237–248, John Benjamins, Amsterdam/Philadelphia <http://www.opensubtitles.org/>
- Kay, Paul, & Fillmore, Charles. 1999. Constructions and linguistic generalizations: The what’s X doing Y? construction. *Language* 75(1): 1–33. doi:10.2307/417472
- Korotkova, Natalia. 2016. *Heterogeneity and Uniformity in the Evidential Domain*. PhD Dissertation, University of California, Los Angeles.
- Littell, Patrick, Matthewson, Lisa & Peterson, Tyler. 2010. On the semantics of conjectural questions. In *Evidence from Evidentials*, Tyler Peterson & Uli Sauerland (eds), 89–104. Vancouver: University of British Columbia Working Papers in Linguistics (UBCWPL).
- Meyer, Wulf Uwe, Reisenzein, Rainer & Schützwohl, Achim. 1997. Toward a process analysis of emotions: the Case of Surprise. *Motivation and Emotion* 21(3): 251–274.
doi:10.1023/A:1024422330338
- Miceli, Maria & Castelfranchi, Cristiano. 2015. *Expectancy and Emotion*. Oxford: OUP.

- Michaelis, Laura & Feng, Hanbing. 2015. What is this, sarcastic syntax? *Constructions and Frames* 7(2): 148–180.
- Michaelis, Laura & Knud, Lambrecht. 1996. The exclamative sentence type in English. In *Conceptual Structure, Discourse and Structure*, Adele Goldberg (ed.), 375–389. Stanford CA: CSLI.
- Mihas, Elena. 2014. Expression of information source meanings in Ashéninka Perené (Arawak). In *The Grammar of Knowledge, A Cross-Linguistic Typology*, Alexandra Aikhenvald, Robert Dixon & Malcolm Ward (eds), 209–226 Oxford: OUP. doi:10.1093/acprof:oso/9780198701316.003.0010
- Ortony, Andrew, Clore, Gerald L. & Collins, Allan. 1988. *The Cognitive Structure of Emotions*. Cambridge: CUP. doi:10.1017/CBO9780511571299
- Pesetsky, David. 1987. Wh-*in-Situ*: Movement and unselective binding. In *The Representation of (in)Definiteness*, Eric Reuland & Alice ter Meulen (eds), 98–129. Cambridge MA: The MIT Press.
- Peterson, Tyler. 2017. Grammatical evidentiality and the unprepared mind. In *Expressing and Describing Surprise* [Benjamins Current Topics 92], Celle Agnès & Laure Lansari (eds), 51–89. Amsterdam: John Benjamins.
- Pinker, Steven. 2007. *The Stuff of Thought: Language as a Window into Human Nature*. New York NY: Penguin.
- Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: OUP.
- Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(2): 165–198. doi:10.1515/TL.2007.011
- Reese, Brian. 2007. *Bias in Questions*. PhD dissertation, The University of Texas at Austin.
- Reisenzein, Rainer. 2000. Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion* 14(1): 1–38. doi:10.1080/026999300378978
- Rett, Jessica & Murray, Sarah. 2013. A semantic account of mirative evidentials. In *Proceedings of SALT 23. Proceedings of the 23rd Semantics and Linguistic Theory Conference, held at the University of California, Santa Cruz, May 3–5, 2013*, Todd Snider (ed.), 453–472. <<https://journals.linguisticsociety.org/proceedings/index.php/SALT/issue/view/85>>
- Rocci, Andrea. 2007. Epistemic modality and questions in dialogue. The case of Italian interrogative constructions in the subjunctive mood. In *Tense, Mood and Aspect: Theoretical and descriptive issues*, Louis de Saussure, Jacques Moeschler & Genoveva Puskas (eds), 129–153. Amsterdam: Rodopi.
- Rohde, Hannah. 2006. Rhetorical questions as redundant interrogatives. *San Diego Linguistics Papers*, Vol. 2, 134–168. San Diego CA: University of California.
- San Roque, Lila, Floyd, Simeon & Norcliffe, Elisabeth. 2017. Evidentiality and interrogativity. *Lingua* 186–187: 120–143.
- Scherer, Klaus, Clark-Polner, Elizabeth, Mortillaro, Marcello. 2011. In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology* 46(6): 401–435. doi:10.1080/00207594.2011.626049
- Siemund, Peter. 2017. Interrogative clauses in English and the social economics of questions. *Journal of Pragmatics* 199: 15–32.
- Stein, Nancy L. & Hernandez, Marc W. 2007. Assessing understanding and appraisals during emotional experience. In *Handbook of Emotion Elicitation and Assessment*, James A. Coan & John J. B. Allen (eds), 298–317. Oxford: OUP.

- Stivers, Tanya. 2010. An overview of the question-response system in American English conversation. *Journal of Pragmatics* 42(10): 2620–2626. doi:10.1016/j.pragma.2010.04.002
- Tutin, Agnès. 2017. Surprise routines in scientific writing: a study of French social science articles. In *Expressing and Describing Surprise* [Benjamins Current Topics 92], Agnès Celle & Laure Lansari (eds), 153–172. Amsterdam: John Benjamins.
- Van Dijk, Wilco W. & Zeelenberg, Marcel. 2002. What do we talk about when we talk about disappointment? Distinguishing outcome-related disappointment from person-related disappointment. *Cognition and Emotion* 16(6): 787–807. doi:10.1080/02699930143000563
- Vingerhoets, Ad, Bylsma, Lauren & De Vlam, Cornelis. 2013. Swearing: A biopsychosocial perspective. *Psychological Topics* 22 (2013), 2, 287–304.
- Zaefferer, Dietmar. 2001. Deconstructing a classical classification: A typological look at Searle's concept of illocution type. *Revue Internationale de Philosophie* 2(216): 209–225.
- Zeisler, Bettina. 2017. Don't believe in a paradigm that you haven't manipulated yourself! – Evidentiality, speaker attitude, and admirativity in Ladakhi. *Review of Cognitive Linguistics* 15(2): 515–538. doi:10.1075/rcl.15.2.09zei

Non-finiteness, complementation and evidentiality

The Lithuanian *Accusativus cum Participio* in a cross-linguistic perspective

Aurelija Usonienė and Nigel Vincent
Vilnius University / University of Manchester

The relatedness of non-finite constructions and evidentiality has been observed in various European languages. Passive matrix verbs plus infinitive in English, the corresponding though less productive pattern in Dutch, reportive passives in Danish, and evidential participial constructions in Lithuanian have all received attention in the literature. We continue this line of investigation, focussing on the *Accusativus cum Participio*, found in contemporary Lithuanian only with verbs of communication, cognition and perception. Our quantitative and qualitative corpus-based analysis investigates its distribution in different types of discourse and provides evidence to support the claim that the use of the non-agreeing ‘be’-participle is obligatory because it marks a proposition. We compare our account with similar uses of non-finite verb in other languages.

Keywords: Lithuanian, accusative plus participle, proposition, finiteness, Latin

1. Introduction

Evidentiality as a subcategory of the conceptual domain of epistemicity is defined “in terms of the notion of source of information, evidence, justification or [...] **epistemic justification**” (Boye 2012: 2). The three notions are there regarded as equivalent but for present purposes the label justification is preferable because it is related to a more general notion of justificatory support and “evidential meanings represent different types of epistemic justification [...] for a proposition” (Boye 2012: 19). Following Willet (1988) and many others, we distinguish between two types of evidential qualification: direct and indirect, with the latter in turn being subdivided into reportive and inferential. Evidential meanings (no matter whether

they are semantic or pragmatic) are expressed cross-linguistically in a variety of ways, both grammatical and lexical, as discussed by, among others, Aikhenvald (2004), Wiemer & Stathi (2010) and Boye (2012). One recurrent type is the use of constructions built around non-finite forms such as infinitives and participles. The common use of passive matrix verbs with the infinitive (the so-called *Nominativus cum Infinitivo* – NcI) in English (Noël 2001), and the corresponding though much less productive pattern in Dutch (Noël & Coleman 2010), and reportive passives with infinitive complements in Danish (Boye & Harder 2009; Ørsnes 2011) have all received attention in the literature. Consider the examples in (1) and (2) which illustrate their use in English and Danish:

- (1) English (Noël & Coleman 2010: 160):
 AMERICAN ring doughnuts from The Delicious Donut Co are made from a flour which
 is said to give them a light, fluffy, and non-greasy consistency.
 (BNC A0C 1141)
- (2) Danish (Boye & Harder 2009: 20):
Han siges/menes at være rig.
 He say.PRS.PASS/hold.PRS.PASS to be.INF rich
 ‘He is said/held to be rich.’

There are no NcI constructions in Lithuanian. Moreover, the non-finite (participial and infinitival) forms which traditionally have been regarded as grammatical realizations of indirect evidentiality in the Baltic and Finnic languages (Gronemeyer 1997; Wälchli 2000; Wiemer 2006; Holvoet 2007; Lavine 2010) are neither NcPs nor AcPs; instead they are non-finite forms used as the main predication of the sentence, as in (3) and (4):

- (3) *Profesorius pasakė, kad mano priėmimui astronomijos katedros laborantu priedaraujja dekanas X. Jis pavadinęš mane „nestropiu“ ir sakęš, kad aš ne toks jau geras studentas [...]*
 (CorALit)
 ‘The professor said that the dean was against my appointment as an assistant at the Department of Astronomy. He called (‘call’ AGR.PST.AP.SG.M.NOM) me “negligent” and said (‘say’ AGR.PST.AP.SG.M.NOM) that I was not a good student [...].’
- (4) *Vyruko būta liekno – paspruko pro kaminą.*
 guy.GEN be.NAGR.PST.PP slim.GEN escape.PST.3 through chimney.ACC
 ‘The guy was obviously slim – (he) escaped through the chimney.’ (CCLL)

The first construction with agreeing active participles as illustrated in (3) is used to express reportive evidentiality, while the one with a non-agreeing *-ta* participle

exemplified in (4) is said to involve inferential evidentiality.¹ Recent corpus-based research has shown that the latter, that is to say, the construction with non-agreeing passive *-ma/-ta* participles, is actually no longer in use in contemporary Lithuanian (Usonienė & Šinkūnienė 2017). The present chapter, however, is neither about the latter nor the former construction but is devoted to the analysis of a less studied lexical-grammatical strategy of evidentiality in Lithuanian, namely the construction traditionally called *Accusativus cum Participio* (AcP). Although there are studies devoted to the detailed analysis of the morphosyntactic properties of the various participial constructions (Ambrasas 1990; Arkadiev 2012), to date no corpus-based research has been carried out on the meaning (semantic and pragmatic) of the sentences containing matrix verbs with AcP complements. In contemporary Lithuanian, as in other languages, this construction with agreeing and non-agreeing participles has been found to be used only in the complementation of verbs of communication, cognition and perception (CCP), for instance:

a. AcP with non-agreeing ‘be’ participle (5) or that of lexical verbs (6)

- (5) [...] *žmonės tvirtino jį esant gerą ir dievobaimingą.* (CCLL-news)
 people.NOM claim.PST.3 he.ACC be.NAGR.PRS.AP good.ACC and
 god-fearing.ACC
 ‘People claimed him to be good and god-fearing.’

- (6) [...] *iš jų [...] švelnių garsų suprato juos šypsantis.* (CCLL-fict)
 from their soft sounds understand.PST.3 they.ACC
 smile.NAGR.PRS.AP.REFL
 ‘From their soft sounds they understood/assumed them to be smiling.’

b. AcP with agreeing participles

- (7) [...] *Joris matė Gvildą, einančią per aguonų lauką.* (CCLL-FICT)
 Joris.NOM see.PST.3 Gvilda.ACC go.AP.PRS.SG.F.ACC through
 poppies.GEN field.ACC
 ‘Joris saw Gvilda walking across the poppy field.’

1. Some authors regard *-ma/-ta* evidentials as markers of oblique or evidential mood in contemporary Lithuanian (Spraukienė, Razanovaitė & Jasionytė 2015).

[...] diese Konstruktion (sc. AcP) findet sich zu allen Zeiten im Austausch mit dem A.c.I., jedoch mit dem Unterschied, daß bei der Partizipialkonstruktion der Nachdruck auf der sinnlichen Perzeption einer im Verlauf begriffenen Handlung oder eines Zustands liegt, während beim A.c.I. mehr der Inhalt der Verbalhandlung unter Zurücktreten der Sinneswahrnehmung betont wird.²

Our analysis will focus on the AcP with a non-agreeing active participle of the verb *būti* ‘be’, namely its three tense forms *esant*-PRESENT, *buvus*-PAST and *būsiant*-FUTURE followed by an NP or AP casemarked in the accusative (as in 5 above) or occasionally in the instrumental (AcP-be.NAGR + NP/AP_{ACC/INS}) as in (13):

- (13) [...] *mažinimą iki 40–50 proc. ar pan. mano*
 decrease.SG.M.ACC to 40–50 percent or similar believe.PRS.3
esant suprantamu. (CorALit)
 be.NAGR.PRS.AP understandable.SG.M.INS
 ‘A decrease by 40–50 percent is believed to be understandable.’

The purpose of this corpus-based study is to provide evidence that the non-agreeing ‘be’-participle in the given AcP is a marker both of a proposition and of the indirectness of experience encoded in that proposition, as in (14):

- (14) *Iš šio pranešimo matome baltų tautų mitologijas*
 from this paper see.PRS.PL.1 Baltic.GEN peoples.GEN mythologies.ACC
esant ne tik giminiškas [...], bet ir labai artimas seniesiems
 be.NAGR.PRS.AP not only related.ACC but also very close to the old
indoeuropiečių tikėjimams. (CCLL-acad)
 Indo-European beliefs
 ‘From this paper we can see (= assume) Baltic mythologies to be not only related but also very close to the Indo-European beliefs.’

The verb *matyti* ‘see’ does not code visual perception (direct/immediate perception); it is used as a cognition verb denoting indirect or mental perception (qualificational meaning) (e.g., Woodcock 1959; Borkin 1973; Tasmowski 1989; Verspoor 1996, 2000; Felser 1998; Dik & Hengeveld 1991; Duffley 1992; Nuyts 2001). Its non-finite (participial) complement clause denotes a proposition (‘Baltic mythologies being related’). The function of the matrix verb is to mark the source of information the author has for making the given assertion, namely inference based on perceptual knowledge. Thus, when the AcP with a non-agreeing ‘be’-participle

2. ‘[...] this construction is found in all periods as an alternative to the AcI, albeit with the difference that the participial construction puts the emphasis on the sensory perception of an ongoing action or state, while with the AcI it is more the content of the verbal action which is stressed and which takes priority over the sense perception.’ (our translation)

is used in the complementation of perception verbs the meaning of the matrix verb can be extended.

One of the basic properties required if an expression is to have an evidential reading is that it should have propositional scope (Boye 2012: 215). Hence, it is only in those instances when the complementation of the matrix verb is clausal that we can expect the evidential meaning to be expressed. However, a clause can denote both a state-of-affairs and a proposition.

The aim of our analysis is to show that it is only the lexical-grammatical configuration under analysis, namely CCP verbs complemented by AcPs denoting a proposition that can be regarded as expressing an evidential meaning. As will be dealt with in the analysis section, this meaning is always characteristic of CC verbs complemented by the AcP with non-agreeing participles, which is not always the case with the AcP containing (non-)agreeing participles when used in the complementation of perception verbs.

We suggest distinguishing between two sub-types of AcPs with (non-)agreeing participles in Lithuanian, namely an AcP denoting a second-order entity, i.e. a state-of-affairs and a third-order entity which is a proposition. Disambiguation between the two of them can be made by Boye's (2012: 215) test which states that "when epistemic expressions have as their semantic scope a clause which is ambiguous with respect to the contrast between state-of-affairs and proposition, they disambiguate it as propositional". When the matrix verb *matyti* 'see' has no evidential extension, in other words when it denotes direct perception, it is complemented by an AcP denoting a state-of-affairs, as in (15):

- (15) *Matau vaikus*_{ACC.PL} *žaidžiančius*_{AP.PRS.ACC.PL} / *žaidžiant*_{NAGR.PRS} *sode*_{LOC}.
'I can see children playing in the garden.'

When an AcP denotes a proposition the matrix verb has an evidential meaning (cf. Boye 2012: 218). This meaning is most common with a predicative (Wiemer 2010: 87–90), an impersonal form of a perception verb (e.g. *matyti* 'see'_{INF}, *justi* 'feel'_{INF}, *girdėti* 'hear'_{INF}, *matos(i)* 'see'_{PRS.3.REFL}) used as a predicate in the matrix clause. In these cases, the evidential meaning conveyed can be direct and indirect as illustrated in the examples below:

- (16) Indirect evidential
Ir visiems *matos*, *valstybę ir Bažnyčią* *nė*
 and everybody.DAT see.PRS.3.REFL state.ACC and church.ACC NEG
negalėjus *a. a. Petru* *Vileišiui* *rasti tinkamesnės*
 NEG.can.NAGR.PST.AP late Petras.DAT Vileišis.DAT find.INF better.GEN
vietos [...] (CCLL-non-fict)
 place.GEN
 'It is seen (obvious) to everybody that the state and the Church could not have found a better place for the late Petras Vileišis'

- (17) Direct evidential
 [...] *jau matyti šviečiant ugnele.* (CCLL-fict)
 already see.INF glitter.NAGR.PRS.AP fire.ACC
 ‘A fire/light can already be seen glittering.’

As *that*-clauses are always propositional, perception verbs in the matrix clause have indirect evidential reading. In (18), the matrix clause with the predicative *girdėti* ‘hear.INF’ has a reportive meaning. It is not used to code indirect speech, it codes hearsay as evidenced by the fact that it has an empty subject and the source of knowledge is not specified.

- (18) *Girdėti, kad [...] nebenori grįžti į namus.*
 hear.INF COMP neg.want.PRS.SG.2 return.INF PRP home.ACC
 ‘Word has it you do not want to come back home.’
Iš Tavęs to galima buvo ir tikėtis. (CCLL-news)
 ‘One could expect you to behave like this.’

2. Data and method

This corpus-based study makes use of both quantitative and qualitative methods of analysis to find out what matrix verbs are most commonly complemented by an AcP in authentic language use and what evidential meanings can be conveyed. Distribution of this construction across different discourse types is also very important. All the data for the analysis have been compiled from two monolingual corpora: the Corpus of Academic Lithuanian (*Corpus Academicum Lithuanicum* CorALit – <<http://coralit.lt/>>) and the news, non-fiction and fiction sub-corpora of the Corpus of the Contemporary Lithuanian Language (CCLL – <<http://tekstynas.vdu.lt/tekstynas/>>). The former is a specialized synchronic corpus of written Lithuanian which includes academic texts published in the period 1999–2009 and represents the language of the main fields of study and research developed in Lithuania, namely biomedical sciences (B), humanities (H), social sciences (S), physical (P), and technological sciences (T). Its size is about 9 million words (8,670,613). The basic text types included are monographs, textbooks, manuals, research articles, review articles, book reviews, and abstracts. The size of the fiction sub-corpus of the Corpus of the Contemporary Lithuanian Language (CCLL-fict) is 15,765,554 words and it contains both original texts and translations from various languages. The size of the sub-corpus of news (CCLL-news) is 86,497,837 and that of the non-fiction sub-corpus is 19,355,341 words.

As the sizes of all the (sub-)corpora are different, when it is necessary to compare the number of occurrences of the data contrasted, raw frequency numbers have been normalized per 1 million words (pmw) to make the comparison

statistically valid. None of the corpora used are annotated, therefore the automatic search was possible only for certain key word forms, for instance: *esant* ‘be’NAGR.PRS.AP, *žinome* ‘know’PRS.PL.1, *mačiau* ‘see’PST.SG.1, while all the quantitative and qualitative analysis of their concordances was carried out manually. For instance when analyzing the concordance of the non-agreeing active participle *esant*, a distinction had to be made between two basic types of its use, namely its most frequent use in the adverbial clause (19) and in the AcP (20):

- (19) [...] *vaikas merkiiasi esant ryškiai šviesai*[...].
 child.NOM close-eyes.PRS.3 be.NAGR.PRS.AP bright.DAT light.DAT
 ‘A child closes its eyes when the light is bright.’ (CorALit-S)
- (20) [...] *teigė raštą esant dieviškos kilmės* [...].
 claim.PST.3 script.ACC be.NAGR.PRS.AP divine.GEN origin.GEN
 ‘It [tradition] claims the script to be of divine origin.’ (CorALit-H)

3. Types of participles and the AcP construction

There are two sub-types of the AcP construction in Lithuanian: one of them takes an agreeing participle (AGR.PTCP) and the other one takes a non-agreeing participle (NAGR.PTCP):

- (21) **AcP with agreeing participle**
 [...] *ji dažnai [...] matydavo jį einantį pro savo ar*
 she often see.PST.3 he.ACC go.PRS.AP.SG.M.ACC past her or
Agotos namus. (CCLL-fict)
 Agatha.GEN home.ACC
 ‘She often used to see him going past her own or Agatha’s house.’
- (22) **AcP with non-agreeing participle**
Po kelių minučių pamatė kažką einant
 after a few minutes.GEN see.PST.3 somebody.ACC go.NAGR.PRS.AP
jo link. (CCLL-fict)
 he.GEN towards
 ‘In a few minutes he saw somebody going towards him.’

The constructions in Example (21) as well as (22) are used as predicates of non-finite clauses and denote situations that are directly perceivable. However, as a rule, agreeing participles mainly function attributively as heads of clauses modifying the grammatical direct object casemarked in the accusative, as illustrated in (23):

- (23) *Cintija žinojo visus būdus padedančius atrodyti jaunesnei.* (CCLL-non-fict)
 Lit. ‘Cynthia knew all ways helping to look younger.’
 ‘Cynthia knew all the means which can help one look younger.’

Thus a distinction needs to be made between AcPs with agreeing participles and modifying participle clauses, though it is not always easy to detect the difference. While pre-modification is a very frequent position for the attributive use in Lithuanian, post-modification is also very common. Moreover, as Lithuanian is a language with relatively free word order, an agreeing participle of the complement clause can precede the clausal subject (NP in the accusative case), and in consequence it is not always easy to distinguish between complementation (AcP) and post-modification (relative participle clauses or participle modifiers). This ambiguity is particularly evident in the case of perception predicates such as *matyti* ‘see’ and *girdėti* ‘hear’. However, the contrast is evident in Examples (24) and (25); in (24) there is a clear case of post-modification (‘a cuckoo which talks with his mother’s voice’):

- (24) [...] *kartą eidamas giria jis išgirsta gegutę,*
 ‘Once walking in a forest he hears’ cuckoo.SG.F.ACC
kalbančią motinos balsu. (CCLL-fict)
 talk.AGR.PRS.AP.SG.F.ACC mother.GEN voice.INS
 ‘Once walking in a forest he hears a cuckoo talking with his mother’s voice.’

In (25), the agreeing participle *kalbančią* ‘talking.SG.F.ACC’ is preposed to the NP ‘my mother’; rather than a modifier it functions as the predicate of the complement clause:

- (25) *Tarytum iš-girdau kalbančią savo motiną.*
 as if PFV-hear.PST.SG.1 talk.AGR.PRS.AP.SG.F.ACC my mother.ACC
 ‘I seem to have heard my mother talking.’ (CCLL-fict)

Example (26) on the other hand is ambiguous: it can either refer (a) to a talking parrot or a parrot which could talk, which is a case of modification, or (b) to the perception of the event of parrot-talking, as in the following English instance with an AcI as a complement clause of the verbs ‘see’ or ‘hear’: ‘Many bird owners first got into aviculture after seeing (or hearing) a parrot speak or sing to them’ (<https://www.thespruce.com/top-talking-bird-species-390534>).

- (26) [...] *atvykęs į Ameriką, išgirdo kalbančią papūgą.*
 ‘Having come to America he heard’ talk.AGR.PRS.AP.ACC parrot.ACC
 (CCLL-news)
- i. ‘Having come to America he heard a talking parrot.’
 - ii. ‘Having come to America he heard a parrot talking.’

The next pair of examples have been chosen to show that the two possible readings are very much context-dependent. In (27), the AcP denotes a dynamic situation (process of speaking a language) the perception of which causes surprise, while in (28), a post-modifying participle phrase is used to specify the exact kind of woman the author wants Adome to be:

- (27) *Japonai šiek tiek nustemba, išgirdę mane*
 ‘The Japanese get surprised when they hear me speaking their language.’
kalbančią jų kalba. (CCLL-news)
 speak.AGR.PRS.AP.SG.F.ACC they.GEN language.INS
- (28) [...] *trokštu pamatyti tikrąją Adomę, mylimą moterį [...],*
 wish.PRS.SG.1 see.INF real.ACC Adome.ACC beloved.ACC woman.ACC
kalbančią paprastais žodžiais,
 speak.AGR.PRS.AP.SG.F.ACC plain.INS words.INS
glostančią mano veidą. (CCLL-fict)
 caress.AGR.PRS.AP.SG.F.ACC my.GEN face.ACC
 ‘I wish to see the real Adome, a beloved woman, who speaks plain words, who caresses my face.’

It is only the AcP construction (both with agreeing and non-agreeing participles) used in the complementation of CCP verbs that can convey various evidential qualifications: reportive and inferential. The basic semantic difference between the sentences containing AcPs with agreeing and non-agreeing participles is the specific type of evidential meaning which is directly dependent on the semantics of the matrix verb. In view of the syntax-semantics interaction at work (matrix verbs and AcPs with the two sub-types of participles) it is only natural to expect constructions with communication and hearing verbs to express a reportive qualification of the content of the situation described by a non-finite complement, as in (29) and (30):

- (29) *Finis Polonia – žodžiai, kuriuos lyg kažkas girdėjo*
 ‘Finis Polonia – the words which as if somebody’ hear.PST.3
ištariant T. Kosciuską po [...]
 utter.NAGR.PRS.AP Kościuszko.ACC after
 ‘the loss of the battle to the Russian army’
pralaimėjimo rusų kariuomenei [...] (CCLL-news)
 ‘Finis Polonia are the words which somebody seems to have heard uttered by Tadeusz Kościuszko after the loss of the battle to the Russian army.’
- (30) *Paskutinėje dramoje moters gyvenimo prasmę Ibsenas*
 ‘In his last drama, woman.GEN life.GEN sense.ACC Ibsen.NOM
teigia esant motinystę. (CCLL-non-fict-acad)
 claim.PRS.3 be.NAGR.PRS.AP maternity.ACC
 ‘In his last drama, Ibsen claims the meaning of woman’s life to be maternity.’

Inferential qualification is most common with cognition and perception verbs as in (31) where the context contains an overt indication of the perceptual knowledge (non-visual general sensation) on the basis of which an inference has been made:

- (31) *Jos žingsnį, balsą, išvaizdą ir dar kažką, ko nemačiau, bet jaučiau ją turint. Kažkokį būdo bruožą[...].*
 feel.PST.SG.1 she.ACC have.NAGR.PRS.AP some.ACC character.GEN feature.ACC
 I felt she had (it). Some characteristic feature.' (CCLL-fict)

The semantic difference between the use of agreeing and non-agreeing participles in the AcP (as will be discussed in more detail in Section 4) is strikingly consistent with the observations by Woodcock and Leumann, Hofmann & Szantyr, quoted above (pp. 240–1), concerning the different meanings of the AcP and AcI in Latin and the contrast they make between the interpretation of the situation described by the complement clause as direct vs. indirect or mental/intellectual.

In what follows the analysis of the data collected will be carried out in the light of the synsemantic features discussed in this section. Our aim is to show what evidential meanings are expressed and what the role of syntax-semantics interaction is by focusing on the non-finite participial construction AcP in the complementation of CCP verbs. Thus, we will have a close look at the matrix verbs with an AcP in complementation; the text types this construction occurs; the types of participles (agreeing vs. non-agreeing) that can be used in the construction.

4. Results and discussion

The findings of our quantitative analysis show that the AcP with agreeing and non-agreeing participles is only found in written Lithuanian and mainly in fiction and academic discourse within humanities and social sciences. No occurrences of this construction were found in the sub-corpus of spoken Lithuanian in CCLL.³ The most frequent form of the non-agreeing participle of *būti* 'be' in the AcP is the present tense form *esant*, most commonly in fiction, as shown in Table 1.

Table 1. Distribution of AcP with non-agreeing 'be' participles in fiction, journalistic and academic discourse

| <i>f/pmw</i> | <i>esant</i> 'be'NAGR.PRS.AP | <i>buvus</i> 'be'NAGR.PST.AP | <i>būsiant</i> 'be'NAGR.FUT.AP |
|--------------|------------------------------|------------------------------|--------------------------------|
| CCLL-fict | 35.8 | 4.4 | 1.5 |
| CorALit | 17.1 | 7.8 | 0.23 |
| CCLL-news | 7.5 | 4.0 | 0.5 |

3. There was only one example of AcP with the non-agreeing 'be' participle (*esant*) found in the spoken sub-corpus of CCLL and it did not occur in the context of casual conversation but rather is an instance of formal spoken Lithuanian occurring in a prepared speech or interview.

As can be seen in the table, the AcP with the non-agreeing ‘be’ participle *esant* is twice as frequent in fiction as in academic discourse. However, this construction was found only in the humanities and social sciences sub-corpora and its normalized frequency in these two major academic disciplines is 41.6 *f/pm.w*. Thus, we can conclude that this construction is in fact more common in academic discourse in the relevant fields than in fiction.

The most frequent governing matrix verbs are those of cognition and communication: *manyti* ‘think-that’ constitutes 47% and *teigti/tvirtinti* ‘state/claim’-type verbs 17% of the occurrences of all the matrix verbs used in the concordance of *esant* in CorALit as in (32):

- (32) [...] *užsibrėžiau tikslą [...] nušviesti lietuvių mitologiją, kurią
 ‘My purpose was to cast light on the Lithuanian mythology which.ACC
 manau esant [...] istorijos pamatu.
 think-that.PRS.SG.1 be.NAGR.PRS.AP history.GEN foundation.INS
 which I believe to be the essence of history.’ (CorALit-H)*

The least frequent matrix verbs are perception verbs. We have found only two occurrences of the verb *matyti* ‘see’; one in the humanities text in the corpus of academic Lithuanian and the other one in news discourse (CCLL). In both cases it is a predicative *matyti* ‘see.INF’ and the meaning conveyed is that of indirect perception with an extension to inferential evidentiality as shown in (33):

- (33) *Iš piešinių matyti dailininką buvus geru
 from drawings.GEN see.INF painter.ACC be.NAGR.PST.AP good.INS
 piešėju [...]. (CCLL-news)
 drawer.INS
 ‘From the drawings the painter can be seen to have been a good drawer.’*

Of special significance is the fact that the AcP with agreeing participles was not found in the complementation of communication and cognition verbs. Although in studies guided by the intuitions of native informants, non-agreeing participles have been regarded as substitutable for the agreeing ones in the complementation of cognition verbs, in actual use it is only the AcP with non-agreeing participles that are attested in the complementation of CC matrix verbs. This in turn implies that the situation denoted by the AcP with a non-agreeing participle cannot be regarded as an event happening in time which can be experienced directly by our senses but rather as a fact which is more abstract and a-temporal, as in the following examples:

- (34) *Kronininkas P. Dusburgietis teigia prūsus*
 ‘Chronicer Peter von Dusburg claim.PRS.3 Prussians.ACC
turėjus šventas giraites, laukus ir vandenis [...]
 have.NAGR.PST.AP sacred.ACC woods.ACC field.ACC and waters.ACC
 ‘The chronicler Peter von Dusburg claims the Prussians to have had sacred
 woods, fields and waters.’ (CCLL-acad)
- (35) *Jos akcentas išdavė ją esant iš Rytų.*
 her accent.NOM betray.PST.3 she.ACC be.NAGR.PRS.AP from East.GEN
 ‘Her accent revealed her to be from the East.’ (CCLL-fict)

It is important to note that in Latin, by contrast, the usual pattern for complementation of CC verbs is AcI.⁴ This parallel suggests that the Lithuanian AcP with non-agreeing participles can be regarded as corresponding to the AcI in Latin, as in (36a) beside its Lithuanian translation in (36b):

- (36) a. *aio te Aiacida Romanos vincere*
 say.PRS.SG.1 you.ACC son of Aeacus.VOC Roman.ACC.PL conquer.INF
posse
 be able.INF
 ‘I say that you, son of Aeacus, can conquer the Romans’ (Ennius *Ann* 174)
- b. *sakau tave, Ajakidai, galint nugalėti*
 say.PRS.SG.1 you.ACC Aiacida.VOC be able.NAGR.PRS.AP conquer.INF
Romėnus
 Roman.ACC.PL

The situation denoted by the AcP with non-agreeing ‘be’ participles is not directly accessible to the human senses; it is an abstract entity which is comparable to the entities denoted by finite *that*-type complement clauses in other languages, as in (37):

- (37) *Jis nė nenutuokė, kas liko kietajame diske, bet žinojo*
 ‘He had no idea what had been left in the hard disk, but know.PST.3
jį nesant tuščią. (CCLL-fict)
 it.ACC NEG.be.NAGR.PRS.AP empty.ACC
 he knew it not to be empty.’

4. It is worth noting the fact that in these constructions the subject is in the accusative case, which means that when the nucleus of the AcI or AcP is a transitive verb taking a direct object, the latter will also be in the accusative. In such cases, world knowledge might be the only way to determine the correct interpretation of the two nouns in the accusative in examples such as Lithuanian: *žinau Homerą. ACC Iliadą. ACC sukūrus. NAGR. PST* (Mindaugas Strockis, personal communication, May 2017) ‘I know that Homer composed the Iliad’ or Latin *purpureamque uvam. ACC facit albam pampinum. ACC habere* (Lucilius 1224) ‘it (the sun) caused the white vine-shoot to bear purple grapes’. Compare also the examples in (36a and 36b) below.

The AcP in (36) can in fact be substituted for a *kad/jog-* ‘that’ clause with no difference in meaning: [...] *bet žinojo*_{know.PST.3'} *kad*_{COMP} *jis*_{it.M.NOM} *nėra*_{NEG.be.PRS.3} *tuščias*_{empty.NOM} ‘but he knew that it was (is) not empty’. On the other hand the AcP with agreeing participles (found only with perception verbs) denotes a situation which can be experienced directly. These instances can be paraphrased by finite *kaip* ‘how’ complement clauses, as in (38):

- (38) [...] *prieš dvejus metus mačiau tave einančią*
 two years ago see.PST.SG.1 YOU.ACC go.AGR.PRS.AP.SG.F.ACC
per tiltą [...] (CCLL-fict)
 across bridge.ACC
 ‘Two years ago I saw you going across/crossing the bridge.’
 ⇒ [...] *mačiau tave kaipėjai per tiltą.*
 [...] ‘I saw you how you were crossing the bridge.’

The semantics of finite *kaip* ‘how’ complement clauses following perception verbs in Lithuanian is the same as that of *kak* ‘how’ clauses in Russian (Noonan 1987: 131; Aikhenvald 2004: 121). ‘How’ clauses can be contrasted with ‘that’ complement clauses in respect of the semantic feature: direct/immediate vs. indirect/mental perception (Dik & Hengeveld 1991: 241–242).

According to Aikhenvald (2004: 122) “[c]omplement clauses as evidentiality strategies tend to be restricted to verbs of perception and cognition” and these clauses should have “the general complementizer”, that is a *that*-type complementizer, which is the case with *kad* ‘that’ clauses in Lithuanian. This same kind of evidential meaning is characteristic of the AcP with non-agreeing participles whereas the AcP with agreeing participles which denote perceivable entities (state-of-affairs) is not used in the complementation of matrix verbs which have an evidential meaning/extension. The fact that the AcP with non-agreeing participles can denote directly non-perceivable situations should be regarded as a piece of evidence to support our claim that this sub-type of the construction can be considered to be on a par with the AcI in English, a language in which, as convincingly argued by Dik and Hengeveld (1991: 241), knowledge predicates and mental perception constructions “make use of *that* complements and acc-cum-inf complements”.

The lexical meaning of the matrix verb plays an important role for the realization of the evidential meaning since the AcP with communication verbs usually conveys a reportive meaning while cognition and perception verbs can convey both inferential and reportive evidential meaning. Inference can be based on perceptual, conceptual knowledge or hearsay, as instantiated in (39) and (40):

(39) Reportive

J. D. Hoffmannas mini Liubčioje dirbus
 Hoffmann.NOM mention.PRS.3 Liubčia.LOC work.NAGR.PST.AP
liuterony spaustuvininką Joną Langę. (CCLL-acad)
 Lutheran.GEN printer.ACC John.ACC Lange.ACC
 ‘Johann Daniel Hoffmann mentions the Lutheran printer Joannes Langius as
 having worked in Liubčia.’

(40) Hearsay

Žinojo jį kenčiant nuo osteosarkomos ir
 know.PST.3 he.ACC suffer.NAGR.PRS.AP from osteosarcoma.GEN and
paramnezijos. (CCLL-fict)
 paramnesia.GEN
 ‘They knew him to be suffering from osteosarcoma and paramnesia.’

An anonymous reviewer has raised a question regarding the evidential meaning of *mini* ‘mention.PRS.3’ taking an AcP as its complement clause in (39) by comparing it to the English sentence with a verb of saying: ‘John says that Bill has worked in London’. First of all, the latter sentence conveys no evidential qualification because it is an instance of reported speech. Second, *minėti* ‘mention’ in the given example is not a mere verb of saying because it does not denote a specific act of speech. On the contrary it is a verb expressing the author’s personal voice in the narrative argumentation. This voice reflects the author’s commitment towards the propositional content of the assertion coded by the AcP. *Minėti* ‘mention’ is expressive of epistemic commitment regarding the content of the written document (a chronicle, a letter, etc.). In consequence, it would not be legitimate to derive a form of direct speech such as ‘Hoffman says’: “*Johannes Langius worked in Liubčia*” analogous to ‘John says’: “Bill has worked in London” because *mini* ‘mention’_{PRS.3} does not refer here to a specific act of verbal communication performed by a specific person, a Mr. Hoffmann, at a specific moment of history. Third, in Lithuanian, the verb *sakyti* ‘say’ has been found to be extremely rarely complemented by an AcP. Only a few occurrences of this use have been found in the corpora analysed and those were the cases when its meaning has been extended to that of cognition verbs like *find* or *consider*. Moreover, such instances occur in specialised narrative discourse (anthologies of religious tracts) and in texts which were mainly translations from Latin and Greek of various philosophical writings by Aristotle and Plato as in (41):

(41) (CCLL-non-fict-translation)

<http://classics.mit.edu/Plato/cratylus.html>

SOKRATAS. *Ogi, manau, jis turi omeny ne tai, kad aukso giminė gimusi iš aukso, bet kad ji gera ir graži. Šito požymis man yra tai, jog mus_{we.ACC} jisai_{he} sako_{say.PRS.3} esant_{be.NAGR.PRS.AP} plieno gimine_{race.ACC}.*

HERMOGENAS. *Tiesą kalbi.*

SOKRATAS. *Ar nemanai, kad ir apie kokį nors iš dabartinių [žmonių], jei kuris yra geras, jis pasakytų, kad tas priklausęs anai aukso giminei?*

HERMOGENAS. *Panašu, kad taip.*

Soc. What is the inference! Why, I suppose that he means by the golden men, not men literally made of gold, but good and noble; and I am convinced of this, because he further says that we are the iron race.

Her. That is true.

Soc. And do you not suppose that good men of our own day would by him be said to be of golden race?

Her. Very likely.

In short, the construction here is grammatically acceptable but actually not common in authentic language use.

The analysis of the data also shows that the two types of meaning, namely direct perception and indirect evidentiality, can be coded by the AcP containing agreeing or non-agreeing participles of full lexical verbs only when they are complement clauses of perception verbs, as shown in the following examples:

(42) Direct perception

[...] *policijos pareigūnai įtartinus asmenis pamatė einančius gatve.* (CCLL-news)
 police.GEN officers.NOM suspected.ACC persons.ACC see.PST.3
 go.AGR.PRS.PL.M.ACC street.INS
 ‘The police officers saw the suspected persons walking the street.’

(43) Indirect evidential

Girdėjau tave [...] nelankant skaitymo pratybų.
 hear.PST.SG.1 you.ACC NEG.attend.NAGR.PRS.AP reading classes.GEN
 ‘I heard that you are not attending reading classes.’ (CCLL-fict)

(44) Indirect evidential

[...] *istorijos archyve saugomuose 1873 ir 1882 metų planuose matome buvus [...] išplėtotą dvaro sodybos ir parko struktūrą;* (CorALit-H)
 be.NAGR.PST.AP extended.ACC estate.GEN and park.GEN
 structure.ACC
 ‘we can see the structure of the estate and the park to have been an extended one.’

On the other hand, the sub-type with the non-agreeing ‘be’ participle can have only indirect evidential meaning, which is not dependent upon the semantics of the matrix verb as in (45) and (46):

- (45) [...] *kas buvo bloga mūsų civilizacijoje, ką mes patys*
 ‘What was wrong in our civilization, what.ACC we ourselves
matëm esant bloga [...]. (CCLL-news)
 see.PST.PL.1 be.NAGR.PRS.AP bad.NAGR.ADJ
 ‘what we ourselves saw to be bad.’
- (46) *Nors dokumentai skelbia jį esant Juozapą,*
 declare.PRS.3 he.ACC be.NAGR.PRS.AP Joseph.ACC
profesorius šio vardo nemėgsta [...] (www.tekstai.lt/.../)
 ‘Though the documents state him to be Joseph, the professor does not like
 this name.’

The use of the non-agreeing ‘be’ participle is always obligatory in the AcP;⁵ its omission renders the clause grammatically unacceptable, as can be clearly seen when we compare Examples (47) and (48):

- (47) *Girdėjau žmones kalbant ši rudenį*
 hear.PST.SG.1 people.ACC speak.NAGR.PRS.AP this.ACC autumn.ACC
būsiant šiltą, gražų ir saulėtą [...]. (CCLL-news)
 be.NAGR.FUT.AP warm.ACC sunny.ACC and sunny.ACC
 ‘I heard people saying this autumn will be warm, nice, and sunny’
- (48) *Girdėjau žmones kalbant *ši rudenį Ø šiltą, gražų ir saulėtą [...].*
 ‘saying this autumn Ø warm, nice, and sunny’

The results of our analysis allow us to conclude that evidentiality is linked to non-finiteness. The AcP construction with a non-agreeing ‘be’ participle in Lithuanian (AcP-be.NAGR + NP/AP_{ACC/INS}) is a predicative non-finite clausal complement denoting a proposition. A proposition is a third-order entity,⁶ which cannot be accessed directly by human senses, hence AcP-be.NAGR + NP/AP_{ACC/INS} is a conceptual construction. When the non-agreeing AcP construction is used in

5. However, in finite clauses the copula can be omitted: *skelbia, kad jis Ø doras žmogus* ‘claims that he (Ø-is) an honest man’. Moreover, when the matrix verb is *laikyti* ‘consider’ or its synonym, the non-agreeing ‘be’ copula can be also omitted in the AcP when the participle of a transitive verb takes an NP or AP casemarked in the instrumental, e.g.:

- (a) *Laiko jį Ø protingu* ‘consider.PRS.3 he.ACC clever.INS’ – Lit. ‘They consider him (Ø-be.NAGR.PRS.AP) clever.’
 (b) *Mato_{see} save Ø imperatoriumi* ‘project/consider.PRS.3 himself.ACC emperor.INS’ – Lit. ‘He considers himself (Ø-be.NAGR.PRS.AP) emperor.’

6. “third-order entities are unobservable and cannot be said to occur or to be located in time and space.” (Lyons 1977: 445)

complementation, the lexical meaning of the matrix perception verb shifts from coding direct experience of some 1st-/2nd-order entity (grammatical object) to indirect or inference-based mental perception which denotes the author's epistemic stance towards the proposition expressed (3rd-order entity). Thus, its function is to mark the author's source of information for the assertion made (evidential meaning), which, as we have noted, seems to correspond to the claims made in the previous studies of Latin and English (Woodcock 1959; Borkin 1973; Duffley 1992) that complement clauses with full infinitive denote indirect perception which can be considered mental or intellectual. By contrast, the AcP with agreeing participles – especially when used in the complementation of perception verbs – denotes situations that can be experienced directly, and hence combine most naturally with matrix verbs which convey the meaning of direct perception. Thus, the AcP used in the complementation of perception verbs can code both direct perception and indirect evidential meaning. At the same time, the distribution of AcPs with agreeing and non-agreeing participles across different discourse types clearly shows their predominance in written narrative discourse (fiction, academic and journalistic). We have found no occurrences of this construction in spoken Lithuanian.

5. Concluding remarks

In conclusion, it is natural to ask why the morphosyntactic properties and the evidential meanings correlate in the way that they do. In particular, why in a number of languages should non-finite forms be susceptible to this kind of interpretation? And what, if any, is the relevance of the dimension of participial (non-)agreement? Let us take each of these in turn.

It is hard to come up with a single criterion or property which unites all the forms traditionally labelled as non-finite (see the contributions to Nikolaeva 2007 and Anderson 2011: Part III for valuable discussion), but a common feature across many approaches to the question is that non-finite forms lack the deictic properties that can anchor them to particular moments of utterance. This holds true even though, as we have seen, traditional grammar labels forms like Lithuanian *esant*, *buvus* and *būsiant* as 'present', 'past' and 'future' respectively, since what is implied here is not absolute or deictic tense, which links an event to the time of speech, but rather relative tense, which links an event to a reference time within the discourse. Participles of the kind we have been examining here will therefore need to be associated with a governing finite verb if they are to be anchored to a particular moment of speech or writing, and thus they will inevitably derive a part of their interpretation from the properties of that governing predicate. The contrast which

we have seen between the value attributed to the complement of a perception verb and that of a verb of communication is hence entirely to be expected.

What then of the role of agreement? Here Lithuanian provides an interesting and valuable test case. At the outset we compared Lithuanian with English and Latin, which in this respect represent two opposite extremes. English has no system of agreement for non-finite forms and hence no contrast is possible. Latin, on the other hand, treats all its participles as part of the class of adjectives; agreement is therefore required across-the-board and so once again no contrast is possible. In all likelihood, the Latin situation is also the one that obtained in the proto-Indo-European stage (see for example the discussion in Lowe 2015); the innovative language is then Lithuanian, which has lost its agreement features in one constructional context but not in the other. The non-agreeing participle shares distributional features with the Latin infinitive, where agreement is not an option, while the agreeing participle has a parallel with the necessarily agreeing Latin participle. Diachrony can also be relevant at the level of content as evidenced in the study of the Latin AcP by Greco (2013) cited above, in which the author shows how the communicative value of the construction and the registers with which it is associated can change over time. A similar register-related restriction is also to be seen in the limitation of the Lithuanian AcP to written rather than spoken discourse.

If the thoughts that we have developed in this final section are along the right lines, we may conclude first that the detailed cross-linguistic examination of the relations between non-finite forms and the expression of evidentiality is potentially a very fertile terrain for future research, and second that languages like Lithuanian in which diachronic splits have occurred in the marking of non-finite forms offer powerful tools for the exploration of this terrain.

Acknowledgement

The research was partially carried out within the framework of the international project EVI-DISPRAG (Reference: FFI2015-65474P (MINECO/FEDER) funded by the Spanish Ministry of Economy and Competitiveness and European Regional Development Fund. We are very grateful to the three anonymous reviewers and the editors for their constructive criticism and valuable advice, which has been very helpful to us in working on the revision of the paper.

List of abbreviations

| | | | |
|---------|-----------------------------------------|------------|----------------------------|
| 1; 2; 3 | person | M | masculine |
| acad | academic | NAGR | non-agreeing |
| ACC | accusative | Ncl | Nominativus cum Infinitivo |
| AcP | Accusativus cum Participio | NEG | negative |
| AGR | agreeing | news | journalistic discourse |
| AP | active participle | NOM | nominative |
| CCP | communication, cognition, perception | PL | plural |
| COMP | complementizer | PL | Polish |
| DAT | dative | PFV | perfective |
| EN | English | PRP | preposition |
| F | feminine | PRS | present |
| fict | fiction | PST | past |
| FUT | future | PTCP | participle |
| GEN | genitive | <i>pmw</i> | per million words |
| H | humanities | REFL | reflexive |
| LAT | Latin | RU | Russian |
| LV | Latvian | SG | singular |
| | | VOC | vocative |

Data sources

CCLLCorpus of the Contemporary Lithuanian Language – <<http://tekstynas.vdu.lt/>>
 CorALit Corpus of Academic Lithuanian (Corpus Academicum Lithuanicum – <<http://coralit.lt/>>)

References

- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: OUP.
- Ambrazas, Vytautas. 1990. *Lietuvių kalbos dalyvių istorinė sintaksė* [Historical Syntax of Participles in the Lithuanian Language]. Vilnius: Mokslas.
- Anderson, John M. 2011. *The Domain of Syntax*. Oxford: OUP.
- Arkadiev, Peter. 2012. Participial complementation in Lithuanian. In *Clause-linkage in Cross-linguistic Perspective: Data-driven Approaches to Cross-clausal Syntax*, Volker Gast & Holger Diessel (eds), 285–334. Berlin: Mouton de Gruyter. doi:10.1515/9783110280692.285
- Borkin, Ann. 1973. *To be or not to be*. *Papers from the Ninth Regional Meeting Chicago Linguistic Society*, Claudia Corum, T. Cedric Smith-Stark & Ann Weiser (eds), 44–56. Chicago IL: CLS.
- Boye, Kasper & Harder, Peter. 2009. Evidentiality. Linguistic categories and grammaticalization. *Functions of Language* 6(1): 9–43. doi:10.1075/fol.16.1.03boy
- Boye, Kasper. 2012. *Epistemic Meaning. A Crosslinguistic and Functional-cognitive Study*. Berlin: Mouton de Gruyter. doi:10.1515/9783110219036

- Cuzzolin, Pierluigi. 1994. *Sull'origine della costruzione dicere quod: Aspetti sintattici e semantici*. Florence: La nuova Italia.
- Dik, Simon C. & Hengeveld, Kees. 1991. The hierarchical structure of the clause and the typology of perception-verb complements. *Linguistics* 29: 231–259. doi:10.1515/ling.1991.29.2.231
- Duffley, Patrick J. 1992. *The English Infinitive*. London: Longman.
- Felser, Claudia. 1998. Perception and control: A Minimalist analysis of English direct perception complements. *Journal of Linguistics* 34: 351–385. doi:10.1017/S0022226798007075
- Greco, Paolo. 2013. Latin *Accusativus cum Participio*: Syntactic description, evidential values and diachronic development. *Journal of Latin Linguistics* 12: 173–198. doi:10.1515/joll-2013-0010
- Gronemeyer, Claire. 1997. Evidentiality in Lithuanian. *Lund University Working Papers* 46: 93–112.
- Holvoet, Axel. 2007. *Mood and Modality in Baltic*. Krakow: Wydawnictwo uniwersytetu Jagiellonskiego.
- Lavine, E. James. 2010. Mood and a transitivity restriction in Lithuanian: The case of the inferential evidential. *Baltic Linguistics* 1: 115–142.
- Leumann, von Manu, Hofmann, Johann & Szantyr, Anton. 1972. *Lateinische Grammatik*. Munich: Beck.
- Lowe, John M. 2015. *Participles in Rigvedic Sanskrit. The Syntax and Semantics of Adjectival Verb Forms*. Oxford: OUP.
- Lyons, John. 1977. *Semantics*, Vol. 2. Cambridge: CUP. doi:10.1093/acprof:oso/9780198701361.001.0001
- Lyons, John. 1977. *Semantics*. Vol.2. Cambridge: Cambridge University Press.
- Nikolaeva, Irina. Ed. 2007. *Finiteness. Theoretical and Empirical Foundations*. Oxford: OUP.
- Noël, Dik. 2001. The passive matrices of English infinitival complement clauses. Evidentials on the road to auxiliarihood. *Studies in Language* 25: 255–296. doi:10.1075/sl.25.2.04noe
- Noël, Dirk & Coleman, Timothy. 2010. Believe-type raising-to-object and raising-to-subject verbs in English and Dutch. A contrastive investigation in diachronic construction grammar. *International Journal of Corpus Linguistics* 15(2): 157–182. doi:10.1075/ijcl.15.2.02noe
- Noonan, Michael. 1987. Complementation. In *Language Typology and Syntactic Description*. Vol. 2, Timothy Shopen (ed.), 42–140. Cambridge: CUP.
- Nuyts, Jan. 2001. Epistemic Modality, Language, and Conceptualization. *A Cognitive-Pragmatic Perspective* [Human Cognitive Processing 5]. Amsterdam: John Benjamins.
- Ørsnes, Bjarne. 2011. Passives and evidentiality: Danish reportive passives and their equivalents in German. *Acta Linguistica Hafniensia* 43(1): 21–59. doi:10.1080/03740463.2011.585042
- Spraunienė, Birutė, Razanovaitė, Auksė & Jasionytė, Erika. 2015. Solving the puzzle of the Lithuanian passive. In *Voice and Argument Structure in Baltic* [Valency, Argument Realization and Grammatical Relations in Baltic 3], Axel Holvoet & Nicole Nau (eds), 323–365. Amsterdam: John Benjamins. doi:10.1075/vargreb.2.07spr
- Tasmowski, Liliane. 1989. A look at *sembler* + infinitive from different Angles. In *Sentential Complementation and the Lexicon. Studies in Honour of Wim de Geest*, Dany Jaspers, Dany, Wim Klooster, Yvan Putseys & Pieter Seuren (eds), 403–438. Dordrecht: Foris.
- Usonienė, Aurelija & Šinkūnienė, Jolanta. 2017. Potential vs use: Revisiting an evidential participial construction in Lithuanian. In *Evidentiality Revisited: Cognitive Grammar, Functional and Discourse-Pragmatic Perspectives* [Pragmatics & Beyond New Series, 271], Juana I. Marín-Arrese, Gerda Haßler & Marta Carretero (eds), 297–340. Amsterdam: John Benjamins. doi:10.1075/pbns.271.08uso
- Verspoor, Marjolijn. 1996. The story of *-ing*: A subjective perspective. In *The Construal of Space in Language and Thought*, Martin Pütz & René Dirven (eds), 417–454. Berlin: Mouton de Gruyter. doi:10.1515/9783110821611.417

- Verspoor, Marjolijn. 2000. Iconicity in English complement constructions. In *Complementation. Cognitive and Functional Perspectives* [Converging Evidence in Language and Communication Research 1] Kaoru Horie (ed.), 199–225. Amsterdam: John Benjamins.
doi:10.1075/celcr.1.08ver
- Wälchli, Bernhard. 2000. Infinite predication as a marker of evidentiality and modality in the languages of the Baltic Region. *STUF* 53(2): 186–210.
- Wiemer, Björn. 2006. Grammatical evidentiality in Lithuanian (A typological assessment). *Baltistica* 36: 33–49.
- Wiemer, Björn. 2010. Hearsay in European languages: Toward an integrative account of grammatical and lexical marking. In *Linguistic Realization of Evidentiality in European Languages*, Gabriele Diewald & Elena Smirnova, 59–130. Berlin: Mouton de Gruyter.
- Wiemer, Björn & Stathi, Katerina. 2010. The database of evidential markers in European languages. A bird's eye view of the conception of the database (The template and problems hidden beneath it). *STUF* 63(4): 275–289.
- Willett, Thomas. 1988. A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language* 12(1): 51–97. doi:10.1075/sl.12.1.04wil
- Woodcock, Eric Charles. 1959. *A New Latin Syntax*. London: Methuen.

The perfect in Avar and Andi

Cross-linguistic variation among two closely-related East Caucasian languages

Samira Verhees

National Research University, Higher School of Economics Moscow

This chapter deals with perfect forms of the verb in Avar and Andi, two East Caucasian languages. The presence of an ergative agent is shown to be an important parameter in distinguishing resultative constructions from resultative perfects in these languages. This distinction is relevant to determine whether current relevance meanings of the perfect are at all represented in these languages, alongside resultative proper and evidential usages. Based on elicitation as well as corpus data, this study shows that the Avar perfect represents a highly polysemic verb form that combines resultative proper, current relevance and indirect evidentiality, while its Andi counterpart shows a more advanced stage of grammaticalization of the indirect evidential meaning.

Keywords: perfect, evidentiality, resultative, East Caucasian languages, Avar

1. Introduction

The perfect is an elusive verb form that is cross-linguistically characterized by instability and polysemy (Plungian 2016). It is also widely attested among the East Caucasian languages, a family of 29 languages that are indigenous to the Caucasus (according to the traditional classification as presented, for example, in van den Berg 2005: 182). Perfects that express indirect evidentiality are considered an areal feature among genetically and structurally diverse languages spoken in a large area that includes and surrounds the Caucasus (see Plungian 2010 for a concise overview). The present chapter explores the nature of the perfect in Avar and Andi, two languages from the Avar-Andic branch of the East Caucasian language family.

I will use the term ‘perfect’ to refer to language-specific verb forms that may express any of the following universal gram-types: resultative, current relevance and indirect evidentiality, while keeping in mind that in order to qualify as a perfect, the

form should either express current relevance at least in some contexts, or should be able to express both resultative and indirect evidentiality. I use the term universal gram-type following Bybee et al. (1994: 48). In my view, this term refers to a meaning expressed by a grammatical form, which is reflected in its usage. I replaced the notion of anterior employed in Bybee et al with the generalized descriptive label ‘current relevance’, which comprises several more concrete meanings (following Ritz 2012). I consider resultative to be a distinct gram-type from what is often referred to as perfect of result or stative perfect, which I subsume under the umbrella-term current relevance. This decision will be elaborated upon in Section 2. In order to account for the mechanisms that give rise to the polysemy of perfect-like forms, it is necessary to view instances of their usage as particular constructions that can be grouped by the rules that govern them. Relevant parameters for Avar and Andi (and potentially also other languages) include the lexical semantics of the verbs that enter into the construction, as well as the presence of an agent.

I will first give an overview of East Caucasian perfects from a typological point of view (Section 2); I will summarize some relevant theoretical discussions and detail my own approach. In Section 3, I will discuss my data from the languages under discussion, followed by a conclusion in which I summarize the most important findings from a typological perspective.

2. Typology of the perfect and East Caucasian languages

In Lindstedt (2000), current relevance is defined as the cross-linguistically valid prototypical meaning of perfect forms, and some authors implicitly assume this concept to be synonym to the term perfect. Following Ritz (2012), I interpret current relevance as a generalized descriptive label used to refer to a set of more concrete meanings, namely:

1. Perfect of result or stative perfect: denotes a state that obtains as the result of a past event (e.g. ‘He has arrived (and is now here)’).
2. Universal perfect or perfect of persistent situation: a situation held for the duration of a definite time interval before reference time and still persists (‘I have lived here for seven years (and still live here)’).
3. Perfect of recent past or ‘hot news’ perfect – an event occurred shortly before the moment of speech (‘The president has *just* resigned!’).
4. Existential or experiential – an event occurred at least once before reference time (‘I have been to Australia (once in my life)’).

2.1 The problem of the resultative

Perfects often originate from resultative constructions (as noted in Nedjalkov & Jaxontov 1988 and Bybee et al. 1994). Resultative constructions form monovalent, stative predicates (Plungian 2016: 10); they denote states obtaining at reference time which are the result of a past action. This is illustrated by Example (1a): an agent can be introduced here only with the help of the preposition “by”, as in (1b). This distinguishes resultative constructions from the perfect of result (Examples (1c) and (1d)), which retains the argument structure of the lexical verb. Whereas resultative constructions always denote a present state, the perfect of result may denote a state that persists at reference time, but this is not necessarily the case. Therefore, resultative perfects are not compatible with adverbs like “still”, as opposed to resultatives proper, cf. Examples (1a)–(1d):

- (1) a. The door is (still) closed.
 b. The door is closed by me.
 *The door is closed me.
 *The door is still closed by me.
 c. I have closed the door.
 d. The door has closed.
 *The door has still closed.

The situation in many East Caucasian languages is less straightforward, because as a rule, the same existential copula or auxiliary is used in both cases, and the agent is always in the ergative case. The equivalents of the sentences in (1a)–(1d) in Avar thus look as follows:

- (2) a. nuts'a (zegi) qχ'an b-ugo
 door (still) close.ANT.CVB N-be.PRS
 ‘The door is (still) closed.’
 b. di-tsa nuts'a qχ'an b-ugo
 I-ERG door close.ANT.CVB N-be.PRS
 ‘The door is closed by me.’
 *di-tsa nuts'a zegi qχ'an b-ugo
 ‘The door is still closed by me.’
 c. di-tsa nuts'a qχ'an b-ugo
 I-ERG door close.ANT.CVB N-be.PRS
 ‘I have closed the door.’
 d. Identical to (2a).

The construction in Avar consists of an ANTERIOR CONVERB (ANT.CVБ) of the lexical verb and the present tense existential copula *b-ugo*. It combines with the adverb *zegi* ‘still’, except when there is an agent involved (parallel to (1b) in English). In English as well as in Avar, the presence of an agent cancels the resultative reading and thus blocks the use of the adverb.

In literature on East Caucasian, resultative constructions and perfects of result are not distinguished by authors. Both are referred to as “resultative”, which is considered a meaning of the perfect. Current relevance meanings besides perfect of result are not very well-attested in these languages (cf. the division of the Avar perfect by Mallaeva (2007: 198–206) into a “perfect of state” (= resultative) versus a “perfect of action” (= mostly evidential), and also Tatevosov’s comments on the difficulty of eliciting current relevance meanings in other East Caucasian languages (2001: 451–452). It is therefore not quite clear whether these forms can actually be considered to possess the meaning of current relevance that is typical of perfects cross-linguistically.

2.2 Grammaticalization: From resultative to current relevance and indirect evidence

Resultative constructions are mostly restricted to verbal lexemes that signal a change of state, although there are a few exceptions (Nedjalkov 1983: 25). When these constructions expand their usage to verbs that do not denote a change of state, they acquire new grammatical meaning. This is a widely attested grammaticalization pattern, which gives rise to current relevance or indirect evidential meanings, see Dahl (1985; 2000) and Bybee et al. (1994). The indirect evidential meaning is generally assumed to arise as a conversational implicature – resultative constructions emphasize the presence of a result at reference time, so that the event that caused it is backgrounded. As a consequence, the implication may arise that the speaker witnessed only the result of a certain action, from which they inferred that this action took place (Bybee et al. 1994, 96). In a later stage, this primarily inferential form may expand its scope to contexts in which the speaker merely heard about an event from someone else.

According to Bybee et al. (1994: 105), current relevance (or “anterior”) and indirect evidentiality are two distinct possible grammaticalization paths for resultative constructions, as shown in the simplified version of the semantic map by Bybee et al, from Tatevosov (2001: 461).

Tatevosov argued that at least for East Caucasian, current relevance (“anterior” in Figure 1) is a necessary intermediary in the development from a resultative

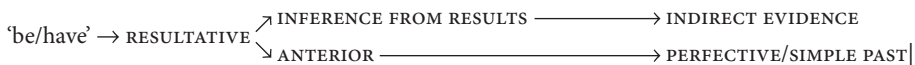


Figure 1. The grammaticalization path(s) of resultative constructions

construction to an indirect evidential marker. The node for anterior on the map in Figure 1 should thus be placed between resultative and inference, instead of on a distinct path (2001: 462). The core of Tatevosov’s argument is that synchronic data from several East Caucasian languages, in which perfects may express resultative, current relevance and indirect evidentiality, force us to assume that one form can develop along two paths simultaneously. This in turn licenses the existence of language-specific forms that can function both as a marked indirect evidential past and as a regular perfective or simple past, for which there is no typological evidence (*ibid.*).

This problem of interpretation can be easily solved by viewing the map as representing three different stages of the grammaticalization process. The first stage is represented by a resultative construction as described earlier in this section – a lexically restricted form that renders a stative situation. The second stage concerns the meanings that arise when this construction expands its usage outside of its usual context (i.e. it starts to combine with non change-of-state verbs). As also described, for example, in Diewald (2006: 4), intermediate stages are characterized by multiple opacity. It is therefore no surprise that at this point, both meanings can be represented by the same form in particular contexts, while other contexts may give rise to ambiguity. In the last stage represented in Figure 1, however, the polysemy of the previous stage has disappeared: either current relevance or indirect evidentiality finally wins the competition for grammaticalized meaning, resulting either in a perfective/simple past, or in an evidential past. The original resultative construction may continue to exist throughout this process.

A common feature among East Caucasian languages is that for particular verbs denoting instantaneously realized events, the resultative is employed to express static situations in the present, i.e. there is no verb ‘to sit’ – the situation ‘I am sitting’ would be rendered with a resultative construction ‘I am sat down’. This is the case in both Avar and Andi, but also in Dargwa (Dargwa branch – Tatevosov 2001) and Rutul (Lezgian branch – Verhees 2017), for example. There is no definite inventory of which verbs this concerns exactly and whether they share this feature across the entire language family, but typically it includes changes of posture and changes in the subject’s physical state ‘to sit/lie down’, ‘to become ill/tired’, ‘to fall asleep’.

2.3 Perfects in narrative sequences and with definite past time reference

A negative criterion for identifying perfects is the ability to be used for the main line of a narrative, because this is not compatible with the current relevance meaning considered central to the perfect, according to Lindstedt (2000: 371). The fact that a perfect can be used as a narrative tense signalling indirect evidential past, however, does not automatically mean that the same form cannot occur with current relevance meaning in another context, as is exemplified by data from Avar presented in this paper: the Avar perfect is commonly used in narratives of events not witnessed by the speaker, while it may also signal current relevance in other contexts (see Sections 3.2–3.4). Another well-known context with which current relevance perfects are considered incompatible, are adverbs with a definite past time reference such as ‘yesterday’, for example:

- (3a) It rained last night.
*It has rained last night.

These criteria also apply to resultative constructions, because they refer to a situation in the present, while indirect evidentials are not subject to such restrictions:

- (3b) *noł* *ts’ad* *ban* *b-ugo*. (PFQ14, Avar)
last_night rain N-fall.ANT.CVB N-be.PRS
‘It rained last night (it appears – everything is wet outside).’

As also noted by Tatevosov, based on examples from Bagvalal (an Andic language closely related to Avar and Andi (2001: 452)), the presence of this type of adverbs forces a perfect to be interpreted as an indirect evidential, whereas otherwise it is ambiguous and can signal either current relevance or indirect evidentiality. In sentences without an agent, a three-way ambiguity arises: if the adverb *noł* ‘last night’ were omitted from Example (3b) for instance, the sentence would have three possible readings: ‘The rain has fallen (it is now lying on the ground / the ground is wet).’ (resultative); ‘It has rained (this might be visible because it is wet outside, or someone might say this because someone else tried to argue to the contrary).’ (current relevance); ‘Apparently, it has rained (the speaker is inferring this from the fact that it is wet outside).’ (indirect evidential).

2.4 The functional distribution of universal gram-types associated with the perfect

Based on the arguments elaborated in the previous paragraphs, the functional distribution of universal gram-types that may co-exist within forms traditionally labelled as perfect in East Caucasian, can be summarized as shown in Table 1.

Table 1. Functional distribution of resultative, current relevance and indirect evidential past

| Typological category | | Perfect | |
|-------------------------|--------------------------------|-----------------------------------------------|----------------------------------------|
| Universal gram-type | Resultative | Current relevance | Indirect evidential past |
| | | Change-of-state verbs | |
| Lexical verb | | Non change-of-state verbs | |
| Meaning | | Event completed in the past (+) | |
| | Resultant state in the present | There may be a resultant state in the present | No resultant state in the present |
| | | | Event was not witnessed by the speaker |
| Durativity | ‘still’ | | |
| Definite time reference | | | ‘yesterday’ |
| Syntax | | Main clause only | |
| | | | Narrative sequence |

It should be noted that current relevance includes several more concrete meanings, as listed in Section 2. The term resultative refers to a resultative construction that meets the criteria listed in the table (i.e. it combines only with verbal lexemes denoting a change of state, there is a resultant state in the present and it can combine with adverbs like ‘still’). It is necessary to draw a sharp distinction between resultative and the perfect of result (the latter belonging to current relevance), in order to establish whether current relevance actually exists in the languages under discussion here. Although current relevance typically includes experiential, in Avar and Andic this function is expressed by the aorist. Perfect-based indirect evidentials are compatible with two types of unwitnessed contexts that are fundamentally different: either the speaker infers that something happened based on some visual evidence (i.e. they did not witness the actual event, but merely an obvious consequence of it), or they heard about it from someone else. Whereas the first context demands a result in the present, the latter can only combine with situations where this is not the case.

3. Avar and Andi

The Avar-Andic branch of East Caucasian is composed of the literary language Avar and its dialects on one hand, and the non-written Andic languages on the other (these are: Akhvakh, Andi, Bagvalal, Botlikh, Chamalal, Godoberi, Karata and Tindi). Avar is one of the official languages of Daghestan: one of the autonomous North Caucasian republics within the Russian Federation, and has traditionally been used as a lingua franca among speakers of Andic languages. Andi is spoken in nine villages near the border with the Chechen republic in north west Daghestan. There are two major dialect-groups: Upper Andi (Andi, Gagatl, Chanko, Gunkha, Ashali, Rikvani, Zilo) and Lower Andi (Muni, Kwankhidatl). Each individual dialect corresponds to a particular village by the same name. Speakers of Andi are taught Avar as “native language” in school, while education is in Russian. Language contact with Chechen does exist, although it is not documented and nowhere near as wide-spread as contact with Avar. For the present paper I chose to limit myself to data from the Rikvani dialect only, because a relatively large corpus of natural texts is available in this dialect and dialectal differentiation in Andi, especially with regards to grammatical semantics, remains an understudied topic.

3.1 Verbal morphology and the perfect

Table 2 shows the paradigm of basic past tense forms of the verb in Avar and Andi. The actual inventories of verb forms featured in these languages are far richer than the simplified paradigm in Table 2, which shows only the perfect, its main rival the aorist and forms that are considered to be derived from them.

Table 2. Basic past tenses in Avar and Andi for the verb ‘to read’

| | | |
|----------------------------|--------------------------|---------------------------------------|
| Avar (ts'alize, 'to read') | Aorist | Perfect |
| | ts'al-ana | ts'al-un b-ugo |
| | read-AOR | read-ANT.CVB N-be.PRS |
| Pluperfect | ts'al-un b-uk'-ana | ts'al-un b-uk'-un b-ugo |
| | read-ANT.CVB N-be-AOR | read-ANT.CVB N-be-ANT.CVB N-be.PRS |
| Imperfect | ts'al-ul-e-b b-uk'-ana | ts'al-ul-e-b b-uk'-un b-ugo |
| | read-PRS-PTCP-N N-be-AOR | read-PRS-PTCP-N N-be-ANT.CVB N-be.PRS |
| Andi (b-al'idu, 'to read') | Aorist | Perfect |
| | b-al'i | b-al'i-d |
| | N-read.AOR | N-read-ANT.CVB(PF) |
| Pluperfect | b-al'i-d b-ik'o | b-al'i-d b-ik'o-d |
| | N-read-ANT.CVB N-be.AOR | N-read-ANT.CVB N-be-ANT.CVB(PF?) |
| Imperfect | b-al'i-rado b-ik'o | b-al'i-rado b-ik'o-d |
| | N-read-PROG N-be.AOR | N-read-PROG N-be-ANT.CVB(PF?) |

In both languages, a limited number of verb stems carry an agreement slot for noun class, which agrees with the absolutive argument.¹ Andi can express number with a changeable stem vowel, whereas Avar replaces the class marker with a plural marker *r-* that is the same for all noun classes. The category of aspect is mostly fused with tense in both languages.

The forms listed in Table 2 are restricted to main clauses; in other clause types, non-finite forms such as converbs and participles are used. In Andi, the forms from Table 2 can be used in declarative as well as interrogative main clauses, whereas Avar makes use of a past participle construction in interrogative clauses. As Table 2 shows, two parallel paradigms are in place in Avar. Each paradigm uses a different form of the auxiliary (aorist vs. perfect), and the corresponding forms are opposed in terms of evidentiality – whereas forms with the perfect auxiliary mark events not witnessed by the speaker, the aorist forms are evidentially neutral.

The perfect itself remains ambiguous, because it can express resultative as well as indirect evidentiality, according to Forker (2018) and Mallaeva (2007). In addition, Mallaeva mentions a “hot news” interpretation, cf. Example (5) and its interpretations elicited by Mallaeva using the TAM questionnaire published in Dahl (1985).

- (5) q'iral w-atʃ-un w-ugo (Mallaeva 2007: 204–205)
king M-arrive-ANT.CVB M-be.PRS
1. ‘The king has arrived (the speaker did not see this, but heard about it).’
 2. ‘The king has arrived (this has just happened and was unexpected).’

For Andi a “past witnessed” (aorist) versus a “past unwitnessed” (perfect) are mentioned in Sulejmanov’s description of the Rikvani dialect (1957). A more recent analysis of the Andi dialect by Kibrik (1985) showed that the form described as “unwitnessed” for this particular dialect may also express a resultative meaning. I decided to make a distinction between aorist and perfect for the Rikvani dialect of Andi analogous to the situation in Avar, following Maisak (2016), who confirmed that resultative semantics also occur with the perfect in Rikvani. It should be noted that the alleged aorist from Andi is considered to mark events witnessed by the speaker in Sulejmanov (1957, Rikvani dialect); Salimov (1968, Gagatl dialect) and Kibrik (1985, Andi dialect), although Maisak (2016, Rikvani dialect) and Tsertsvadze (1965, all dialects except Gunkha) do not mention this opposition. It is not quite clear whether the evidential distinction extends throughout the paradigm parallel to the Avar system, because the usage of periphrastic forms is simply not described.

1. As a citation form I use the verb stem that agrees with neuter singular (= the prefix *b-*, although Andi features several neuter noun classes that each have their own marking).

In Rikvani (as in all other dialects except KwankhidatI) the perfect formally coincides with the anterior converb, because the copula has been dropped. This is part of a general tendency in Andi to omit the copula in the present tense. In Table 2 I chose to make this homonymy of forms explicit by glossing the perfect-suffix *-d* as *-ANT.CVB(PF)*; in the examples throughout this chapter they are glossed according to their function in a specific example, which can be either *ANT.CVB* (ANTERIOR CONVERB) or *PF* (PERFECT).

3.2 Data

The data for this study consists of sentences elicited using the perfect questionnaire (see Dahl 2000); I consulted a speaker of Avar (Khunzakh dialect, which is close to the literary language), and a speaker of Andi (Rikvani dialect). In addition, samples were taken from two corpora: the Rikvani expedition corpus and the Avar Text Corpus.² The Rikvani Expedition Corpus is a collection of spoken texts recorded during fieldwork in 2015.³ It contains a total of 4268 words and is currently in the process of being annotated. I analyzed a sample of 1829 words containing different types of narratives from this corpus (personal memories and local legends), that were recorded with different speakers. From this sample I gathered a total of 42 perfect forms. For Avar I selected over 200 sentences containing perfects from different sub-corpora of the Avar Text Corpus located at <http://baltoslav.eu>, which contains 2,088,693 words in total. From this selection I drew a random sample roughly the same size as the Rikvani sample (53 perfects).⁴ For examples elicited with the questionnaire, the abbreviation PFQ is used to indicate the source. Examples from the Avar corpus are accompanied by a brief reference to a concrete source.⁵

2. I am indebted to my language consultants: Chopan Gaziev for Andi and Magomed Malachiev and Aminat Alibekova for Avar, and also Magomed Abdulkadirovich Magomedov of the Institute of History, Archaeology and Ethnography, DSC of RAS in Makhachkala.

3. This collection is currently unpublished; I would like to thank Michael Daniel of NRU HSE Moscow for sharing this material with me.

4. In the future I would like to take a more quantitative approach to this problem, but the current state of the corpora does not yet allow for such research to be undertaken.

5. More information on the sources from the Avar Text Corpus can be found here: <http://baltoslav.eu/avar/?mova=en#ist>.

3.3 Current relevance

In my data for Avar, two types of current relevance meanings surfaced, as shown in Examples (6) and (7a) elicited with the Perfect Questionnaire.

1. Universal perfect

- (6) do-j j-**alah-un** j-**igo** ʃab-go safat-aʃ. (PFQ54, Avar)
 dem-F F-watch-ANT.CVB F-be.PRS three-NUM hour-PL
 ‘She has watched [TV] for three hours (and she is still watching).’

2. Perfect of result

- (7a) di-tsa gordu **rah-un** **b-ugo.** (PFQ40, Avar)
 I-ERG window open-ANT.CVB N-be.PRS
 ‘I have opened the window (it is now open).’

In Andi, both of these sentences were translated with an aorist, see (7b–c). A translation of (7b) with a perfect was immediately interpreted by the speaker as signalling indirect evidentiality, resulting in the first-person effect that occurs when evidential forms are used with a first person subject. The same construction without an agent in the ergative is a resultative proper (7c).

- (7) b. den ingur **arχon** / ʔden ingur **arχon-d**
 I.ERG window open.AOR / I.ERG window open-PF
 ‘I opened the window.’ / ‘I have opened the window (it appears).’
 c. ingur **arχon-d**
 window open-PF
 ‘The window is open(ed).’

The perfect of result seems to occur only in Avar, as in Example (8), which is from a text about the German activist Clara Zetkin, who is considered to be the founder of International Women’s Day (8th of March). The text states that she has made two mistakes (using the perfect), followed by a clarification that the first mistake she made was to make the 8th of March Women’s day (using the aorist). The surrounding context is rendered in aorist, which confirms that this is not an evidential narrative (as there is no shift in the speaker’s perspective between: “Clara Tsetkin has made two mistakes.” and “The first mistake was to create Women’s Day.”). Because the German woman named Clara Zetkin is in the ergative case, this example cannot be interpreted as a resultative proper.

- (8) klara tsetkin ab-ul-e-j nemtsa-ʃ k'i-go ʃalat'
 Clara Zetkin say-PRS-PTCP-F german-ERG two-NUM mistake
b-itʃi-an **b-ugo.** (Millat 2011/2012)
 N-allow-ANT.CVB N-be.PRS
 ‘The German (woman) by the name of Clara Tsetkin has made two mistakes.’

In my data for Andi, only one example of what may be considered current relevance occurred. According to the speaker, the perfect in (9) was appropriate in this example, because the event was entirely unexpected, even though it was witnessed directly by the speaker.

- (9) χ an w-ugo-d. (PFQ56, Andi)
 king M-arrive-PF
 ‘The king has arrived (it was totally unexpected).’

This type of function is reminiscent of Example (5) from Avar, although no analogous examples showed up in my own data of Avar. The fact that this meaning surfaces with the verb ‘to arrive’ in particular is perhaps telling. Verbs like ‘to come’ or ‘to arrive’ cross-linguistically seem less capable of forming resultative constructions, cf.

- (10) a. He has arrived.
 b. *He is arrived.

3.4 Resultatives and agents

Resultative constructions were attested in both languages, although in Andi they occurred only in a few examples elicited with the Perfect Questionnaire, such as Example (11) below.

- (11) hege-w w-utʃi-d (PFQ30, Andi)
 dem-M M-wake_up-PF
 ‘He has woken up.’

In Avar, they occurred a few times in the Perfect Questionnaire, but also in the corpus sample, as in (12).

- (12) pulemet nuts’ik’alt’a tufmanas-da b-ix-ula-re-b
 machine_gun at_the_door enemy-SUP N-be_visible-PRS-NEG.PTCP-N
 sanaʃatab bak’al-da ʃun b-ugo
 convenient place-SUP put.ANT.CVB N-be.PRS
 ‘At the entrance, in a convenient place out of the enemy’s sight, stands (lit. is put) a machine gun.’ (Mikaʒilov, Sh. 2004. Antsuqχ k’alaqχ sordo)

As mentioned in Section 2.1, in Avar an ergative agent can be introduced to a resultative construction. As a result, the construction becomes less stative and loses its ability to combine with the adverb *zegi* ‘still’. Constructions with or without

an ergative agent in Avar thus seem to form a minimal pair that distinguishes resultative constructions from something similar to a perfect of result. In Andi, the presence of an ergative agent forces an evidential reinterpretation of the sentence, as was demonstrated in Example (7b).

While during elicitation a few instances of resultative constructions occurred in Andi, the corpus data contains none. This discrepancy is due to the content of the corpus: it contains narratives from personal memory and local legends, in which only verb forms with past tense reference are used. The Avar corpus on the other hand, contains journalistic texts and literary writing. The former type of texts is known to frequently feature resultative perfects used as impersonal passive constructions, as was already observed by Forker (2018). Avar does not have a formal device for passive voice that is distinct from the resultative / perfect construction (and neither does Andi) (cf. Examples (1a–d) from English and (2a–d) from Avar), so that this passive-like use should probably be seen as a stylistic device specific of Avar journalistic writing. Literary texts such as novels are less restricted to certain narrative regimes, as Example (13) shows.

- (13) *ko-w* *tʃijas* *ab-una*: «*di-r ganʃal r-il-un*
dem-M person say-PRS I-GEN buffalo.PL PL-become_lost-ANT.CVB
r-ugo. (Saidov M.S. 2004. *Naborshik – Maʒarul proza*)
PL-be.PRS
 ‘That person said: “My buffaloes have disappeared.”’

3.5 Non-witnessed and witnessed events

In Avar as well as Andi, the perfect may be used to express indirect evidential meaning. Elicitation shows, however, that whereas in Avar this is optional (although strongly conventionalized), Andi (at least the dialect of Rikvani) features a consistent opposition of witnessed (aorist) versus unwitnessed (perfect). The perfect questionnaire contains several questions aimed at eliciting the same small narrative viewed from different perspectives, including distant past, recent past, witnessed and unwitnessed past. Speakers of both languages translated the witnessed version with the aorist and the unwitnessed version with the perfect. (14a–b) are the unwitnessed versions of the narrative.

- (14) a. Ha-w ine w-uk'-un w-ugo roxosan, borxi-da
 dem-M go.INF M-be-ANT.CVB M-be.PRS in_the_forest snake-SUP
 t'ad tʃun w-ugo, he-ɬ has-da hat'i-da
 on_top stop-ANT.CVB M-be.PRS dem-ERG dem-SUP foot-SUP
 hantʃ-un b-ugo, ha-s gamatʃ-gi bos-un
 bite-ANT.CVB N-be.PRS dem-ERG rock-ADD take-ANT.CVB
 retʃi-un b-ugo borxi-da he-b χun b-ugo
 throw-ANT.CVB N-be.PRS snake-SUP dem-N die.ANT.CVB N-be.PRS
 'He walked_{PERFECT} in the forest, stepped_{PERFECT} on a snake, it bit_{PERFECT}
 him in the foot, he took_{CONVERB} a rock and threw_{PERFECT} it at the snake. It
 died_{PERFECT}' (PFQ60, Avar)
- b. reʃuɬ w-uʒin-no-guʒa ts'inni-gu-tʃigu berku-l'a
 in_the_forest M-walk-HAB-SIM.CVB know-EMPH-NEG.PF snake-SUP
 tʃek'a b-izi-d hegel-d tʃek'u-tʃu q'ammi-d hegeʃu-d
 leg N-stop-PF dem-ERG leg-AD bite-PF dem.M-ERG
 b-iχi-d=lo hints'o=lo tʃaj-d hege-b b-itʃoli-d
 N-take-ANT.CVB-ADD rock-SBR hit-ANT.CVB dem-N N-die.CAUS-PF
 'While walking_{CONVERB} in the forest, suddenly (lit. not knowing) [he]
 stepped_{PERFECT} on a snake, it bit_{PERFECT} [him], he took_{CONVERB} a rock, hit-
 CONVERB [it] and killed_{PERFECT} it.' (PFQ60, Andi)

In Andi, however, the unwitnessed version can be told using the perfect only – if these perfects are replaced with aorists, the story is automatically interpreted by the speaker as narrating directly witnessed events. In Avar this is not necessarily the case; the aorist does carry some implication that the speaker has directly witnessed the events, but it can be used in a non-witnessed context nonetheless, rendering a more factual account of events. Though the perfect in Avar is polysemic, in a connected narrative only the evidential interpretation is available. The Andi aorist may occur as an unmarked form in statements that do not describe a particular event but present something as a fact, as shown in Example (15).

- (15) hege-b grinʃu-d qχwari (PFQ26, Andi)
 dem-N Greene.M-ERG write.AOR
 'Graham Greene wrote it.'

In the Rikvani corpus, perfects are consistently employed to describe events not witnessed by the speaker. The majority of the examples in the sample are from recordings of local legends, which are unwitnessed events by default.

- (20) *ko-s* *tsogida-l qxuluqxɬfaʒaz-dasa-gi,* *di-dasa ʒadin,* *ts'aq' ʒemer*
 dem-ERG other-PL employee-SUP.EL-ADD I-SUP.EL as very much
ħalt'i t'alab hab-ul-e-b b-uk'-un b-ugo
 work demand make-PRS-PTCP-N N-be-ANT.CVB N-be.PRS
 'He demanded a lot of work from me, as well as from the other workers.'
 (Saidov M.S. 2004. Naborshik – Maʒarul proza)

As opposed to Example (21).

- (21) *habsaʒat,* *di-r emen ʒajib-habuge ʒaze-go ʒatʃogo*
 now I-GEN father forgive_me know.INF-NEG unknowingly
swaka-n w-uk'-un w-ugo dun
 become_tired-ANT.CVB M-be-ANT.CVB M-be.PRS I
 'Now, my father, forgive me, unknowingly it seems I have become tired.'
 (Dinmuhamaw R. 2004. Biduqχ bi)

4. Conclusion

The data gathered for this study show that the perfect in Avar can express resultative (i.e. it can function as a resultative construction, as described in Section 2), as well as two kinds of current relevance meanings (perfect of result and universal perfect) and indirect evidentiality, although the latter seems to be a conventionalized implicature, rather than a grammaticalized meaning. In the Rikvani dialect of Andi, the indirect evidential meaning appears to have grammaticalized, resulting in a consistent opposition between the aorist (witnessed) and the perfect (unwitnessed) that is confirmed by elicitation as well as corpus data. These observations should, however, be corroborated with data from other speakers. The corpus data by itself cannot confirm that this opposition is grammaticalized, as narratives in languages where this is not the case (e.g. Avar) follow the same patterns, using the aorist to relate witnessed events, while the perfect is the default form for unwitnessed events. The degree to which this distribution is an obligatory feature surfaces only during elicitation.

Current relevance meanings are virtually unattested in Andi, with the exception of one example that might be interpreted as a hot news perfect. A fairly clean split seems to have occurred in this language between the resultative construction and the evidential past. Consequently, resultative use is not very frequent, as it is subject to the same restrictions as a resultative proper, which results in a limited distribution in comparison to the perfect of result. When an ergative agent is introduced, the form is reinterpreted as an evidential, whereas in Avar the interpretation would shift to a perfect of result. This confirms that the presence of an ergative agent can

make a meaningful distinction between resultative constructions and other grams, such as current relevance and indirect evidentiality. Due to the limited size of the samples used in the present study, no meaningful conclusions can be drawn with regards to possible links between lexical semantics of verbs and their interpretation. It remains unclear how forms similar to the perfect, such as the pluperfect or the imperfect formed with a perfect form of the auxiliary, fit into this picture. The sample for the current study produced only a handful of examples, which do not allow for any conclusions to be drawn. The nature of their usage, as well as that of corresponding forms with future tense reference left untouched in the present study, seem fruitful topics for future research.

Another mystery yet unsolved is why and how one language does not grammaticalize indirect evidentiality, whereas another closely related language that is heavily influenced by the former does. Although many theories have been proposed over the years to account for the areal contiguity among languages that feature perfect-like forms with evidential meanings, very little has been conclusively proven. In my data I have found no language-internal cues that specifically give rise to an evidential interpretation (comparable, for example, to the lowering of stativity by adding an agent that may pave the way for current relevance).

Abbreviations

| | | | |
|------|---------------|------|------------------|
| AD | ad (locative) | INF | infinitive |
| ADD | additive | M | masculine |
| ANT | anterior | N | neuter |
| AOR | aorist | NEG | negative |
| ATR | attributive | NUM | numeral |
| CAUS | causative | PF | perfect |
| CVB | converb | PL | plural |
| EL | elative | PRS | present |
| ERG | ergative | PTCP | participle |
| F | feminine | QUOT | quotative |
| GEN | genitive | REP | reportative |
| HAB | habitual | SBR | subordinator |
| IN | in locative | SUP | super (locative) |

References

- Bybee, Joan, Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar*. Chicago IL: The University of Chicago Press.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Basil Blackwell.
- Dahl, Östen. 2000. *Tense and Aspect in the Languages of Europe*. Berlin: Mouton de Gruyter. doi:10.1515/9783110197099
- Diewald, Gabriele. 2006. Context types in grammaticalization as constructions. In *Constructions All Over – Case Studies and Theoretical Implications*, Doris Schönefeld (ed.) (Special Vol. 1 of *Constructions*). <<http://www.constructions-online.de/articles/specvoll/>>
- Forker, Diana. 2018. The semantics of evidentiality and epistemic modality in Avar. In *The Semantics of Verbal Categories in Nakh-Daghestanian Languages*, Diana Forker & Timur Maisak (eds). Leiden: Brill.
- Kibrik, Andrej. 1985. Absentiv-aorist und resultativ in der andischen Sprache. In *Sprachen Europas und Asiens*, Heinz Fähnrich (ed.), 55–66. Jena: Friedrich Schiller Universität.
- Lindstedt, Jouko. 2000. The perfect – Aspectual, temporal and evidential. In *Tense and Aspect in the Languages of Europe*, Östen Dahl (ed.), 365–384. Berlin: Mouton de Gruyter.
- Maisak, Timur. 2016. Glagol'naja sistema Rikvaninskogo govora andijskogo jazyka: Formo-obrazovanije (The verbal system of the Rikvani dialect of the Andi language: Conjugation). Ms, *Andi working seminar*. Institute of Linguistics, Russian Academy of Sciences.
- Mallaeva, Zulajxat. 2007. *Glagol avarского jazyka: Struktura, semantika, funkcii* (The verb in Avar: Structure, semantics and functions). Makhachkala: Institut Jazyka, Literatury i Iskusstva imeni Gamzata Cadasy Dagestanskogo Naučnogo Centra Rossijskoj Akademij Nauk (IJaLI).
- Nedyalkov, Vladimir. 1983. *Tipologija rezul'tativnyx konstrukcij* (The typology of resultative constructions). Leningrad: Nauka.
- Nedjalkov, Vladimir & Jaxontov, Sergei. 1988. The typology of resultative constructions. In *Typology of Resultative Constructions* [Typological Studies in Language 12], Vladimir Nedjalkov (ed.), 3–62. Amsterdam: John Benjamins. doi:10.1075/tsl.12.06ned
- Plungian, Vladimir. 2010. Types of verbal evidentiality marking: an overview. In *Linguistic realization of evidentiality in European languages*, Gabriele Diewald & Elena Smirnova (eds), 15–58. Berlin/New York: De Gruyter Mouton.
- Plungian, Vladimir. 2016. K tipologii perfekta v jazykax mira: Predislovije (Towards a typology of the perfect in the world's languages: Introduction). In *Tipologija Perfekta* [Typology of the perfect], Timur Maisak, Vladimir Plungian, Ksenia Semënova (eds), 7–38. St. Petersburg: Nauk.
- Ritz, Marie-Eve. 2012. Perfect tense and aspect. In *The Oxford Handbook of Tense and Aspect*, Robert Binnick (ed.), 881–907. Oxford: OUP.
- Salimov, Khangerej. 1968. *Gagatlinskij govor andijskogo jazyka* (The Gagatl dialect of the Andi language). Makhachkala: IJaLI.
- Sulejmanov, Jakub. 1957. *Grammatičeskij očerok andijskogo jazyka (po dannym govora s. Rikvani)* (Sketch of the grammar of Andi (based on the sub-dialect of the village Rikvani)). PhD dissertation, Institut Jazykoznanija Akademii Nauk Sojuza SSR, Moscow.
- Tatevosov, Sergei. 2001. From resultatives to evidentials: Multiple uses of the perfect in Nakh-Daghestanian languages. *Journal of Pragmatics* 33: 443–464. doi:10.1016/S0378-2166(00)00012-6
- Tsertsvadze, Ilja. 1965. *Andiuri ena* (The Andi language). Tbilisi: Metsnieriaba.

van den Berg, Helma. 2005. The East Caucasian language family. *Lingua* 115: 147–190.

doi:10.1016/j.lingua.2003.06.004

Verhees, Samira. 2017. Èvidencial'nost' i perfekt v rutul'skom jazyke (na materiale govora s. Kina (Evidentiality and the perfect in Rutul (based on the Kina dialect)). In *Èlektronnaja pis'mennost' narodov Rossijskoj Federacii: opyt, problemy, perspektivy* [Electronic writing of the peoples of the Russian Federation: practice, problems and perspectives], Marina S. Fedina (ed.), 228–235. Syktyvkar: GOU VO KRAGSiU. <<http://fu-lab.ru/conference/files>>

The different grammars of event singularisation

A cross-linguistic corpus study

Eric Corre

Université Sorbonne Nouvelle – Paris 3

This chapter is an empirical investigation into the expression of bounded single situations across four languages, based on a parallel corpus (Camus's *The Stranger* and translations into English, Russian, Hungarian). Smith (1991)'s two-component theory of aspect, whereby situation aspect combines with viewpoint aspect to compute the aspectual composition of sentences, is used to highlight cross-linguistic differences. In the original, the French *passé composé* appears as perfective in the sense of Smith (1991) and Klein (1994) while the English simple past is aspectually ambiguous (perfective and imperfective). Russian relies on a morphosyntactic construction (prefix + bare verb) to create perfective verbs, while Hungarian has similar morphosyntactic resources, but no grammatical aspect.

Keywords: aspect, single situations, coercion, prefixation, verb classes

1. Introduction

This chapter is a usage-based investigation into the expression of event singularisation, and deals with notions like telicity, boundedness and perfectivity (Depraetere 1995; Bertinetto 2001; Borer 2005), in particular the expression of single bounded episodes or situations. The terms 'episodes' or 'situations' are meant to encompass both telic and atelic eventualities, as well as punctual and durative ones. The event classes that are left behind are those expressing multiple, repeated events: habits, iterated activities, progressive states. The choice to focus on single situations is arbitrary: it serves as a probe into the investigation of the aspectual systems of several typologically unrelated languages.

Our empirical and comparative approach is based on a small literary corpus (32,400 words), the French novel *L'Étranger* (*The Outsider*) by A. Camus (1942) and three of its published translations (English,¹ Russian, Hungarian). This contrastive

1. The English translation of the novel chosen for the discussion of examples as well as for translations into English of the original French, is Sandra Smith's 2012 *The Outsider* (Penguin Books).

method which compares language data in parallel corpora has been used by several authors: Furko (2014: 182) considers that “finding translation equivalents is a reliable way of mapping individual [items’] functional spectra and can also serve as a heuristic for establishing semantic-pragmatic fields.” Guillemin-Flescher’s (1981) contrastive method acknowledges that translations are linguistically significant if they are systematic. This cross-linguistic discussion relies on Smith’s (1991) model of aspect, where two components, situation and viewpoint aspect, interact to yield the aspectual composition of a given sentence. Viewpoint (or outer) aspect establishes a relation between the assertion time (AT) and the event time (ET). If ET is included in AT, there is an external view of the situation seen as a whole, whatever its actual length; the situation is temporally bounded, producing the perfective viewpoint (PFV). If, on the other hand, AT is included in ET, there is an internal view of the situation; the situation is temporally unbounded, deriving the imperfective viewpoint (ImpV). Viewpoint aspect then composes with Situation aspect to yield different aspectual construals of situations. This general parameter of aspect has often been criticized for being precisely that, too general – it needs to be augmented by pragmatic conventions of use if one is to account for cross-linguistic differences. But the overarching distinction it provides has been handy when it has come to classifying the occurrences in our source text – that it, the French *passé composé*. The present author is aware that other approaches are available, like more typological ones (Bybee & Dahl 1989; Dik 1989; Tournadre 2004; Plungian 2012) and discourse-oriented or constructional approaches (as in Moens & Steedman 1988; de Swart 1998; Michaelis 2004), which consider that tense-aspect markers are the product of different paths of development and constitute “a repertoire of grammatical constructions” (Michaelis 2016) whose function is to alter verb meanings, so-called shift constructions. Coercion is indeed an important phenomenon in the behavior of aspectual forms, and will be called upon to explain the data (cf. Section 2.2).

This chapter carries out a fine-grained examination of translation equivalents of single-situation French *passé composés* (PC) across three different languages, English, Russian, and Hungarian, which all display different features in the domain of tense-aspect marking. The first section deals with French and English which use tensed inflectional forms (PC and simple past tense, henceforth SP): a major question is whether these constructions, in particular the English SP, should be viewed as aspectually neutral (de Swart 1998), that is, not changing the aspectual potential (Croft 2016) of a given eventuality, or instead, whether they are real type-shifters. The second section focuses on Russian and Hungarian which rely heavily on lexico-morphological resources with a systematic opposition between bare (prefixless) and prefixed verbs, thus displaying sensitivity to the *Aktionsart*² properties of the predicate. The central question will be that of grammatical aspect:

2. I use the term *Aktionsart* as a synonym for lexical/actional event class.

Russian is known to have a strongly grammaticized aspectual opposition, but the picture is much less clear for Hungarian. For some authors (e.g. Kiss 2006), situation aspect is grammaticized while viewpoint aspect is expressed by means of correlations. For others (e.g. Kiefer 1994; Csirmaz 2006), Hungarian aspect exemplifies the correctness of the two-component system, making it similar in essence to what we find in Russian. We contend that only a data-oriented study like this one can begin to provide answers.

In the course of the discussion, each language will be analyzed in turn, with the aim of bringing out the main points of similarity and divergence between them.

2. French vs. English

2.1 The French *passé composé* (PC)

The Stranger is notorious for its dominant use of the PC, which is a good candidate for expressing single episodes; its main function is to impose a retrospective final bound on the situation (Gosselin 1996). The PC is actually a perfect construction with the auxiliaries *avoir* ‘have’ or *être* ‘be’ + the past participle of the verb, and it has “perfectivized” over time (Caudal 2015: 199). It is a marked choice in the novel: contrary to the *passé simple*, which is the expected tense form for expressing narrative continuity, the PC is “not authentically narrative; although it can be coerced into expressing narrative progression, in and of itself it does not introduce a temporal order between the situations that are reported” (de Swart & Molendijk 2002: 203). Thus, it contributes to the stylistic effect, as Sartre (1947: 117–118) observed, of “emphasizing the isolation of each sentence unit” in the novel. As stated in the introduction, we adopt Smith’s (1991) or Klein’s (1994) definition of perfective for the French PC: the event time (situation time in Klein) is included in the assertion time (topic time in Klein), and the right boundary is specified at speech time (utterance time in Klein). This accords with Comrie’s (1976: 16) definition of perfectivity, which “indicates the view of a situation as a single whole, without distinction of the separate phases that make up that situation”. For Hopper (1982), Bybee, Perkins & Pagliuca (1994: 55), with perfectives the “situation is viewed as bounded temporally. [It] is the aspect used for narrating sequences of discrete events in which the situation is reported for its own sake”.

In the corpus, out of the total 1754 occurrences of PC, 1710 are perfective in the sense described above; in contemporary and conversational French, it is one of the two main uses of the *passé composé*, called “*aoriste du discours*” in (Benveniste 1966: 249), that is, a substitute for the narrative *passé simple*. But in conversational modes of discourse, the PC may also take on another function, akin to that of the English resultative present perfect (McCawley 1971): this use is generally called

accompli de présent in the literature (Vet 1980); in the source text, we find 23 of those, 18 of which are translated by English present perfects.³ All these forms (the 1710 perfective ‘*aoristes du discours*’ + the 23 ‘*accomplis de présent*’) refer to single bounded situations but with a currently relevant state that endures at Speech Time for the latter. The remaining 23 PC forms have experiential meaning;⁴ they are left out of the analysis in the discussion of French and English because they denote multiple or repeated events. However, because they interact in interesting ways with the so-called factual existential imperfective in Russian (Padučeva 1996, Grønn 2004), occasional reference will be made to them in Section 3.

2.2 The English simple past (SP): Neutral or shifter?

In the English data, 93% of the 1733 occurrences of single- situation PCs are translated by means of a SP form; the two other main constructions (‘started + V-ing’, present perfect) used for the translation of the PC are negligible, accounting for less than 1% each.

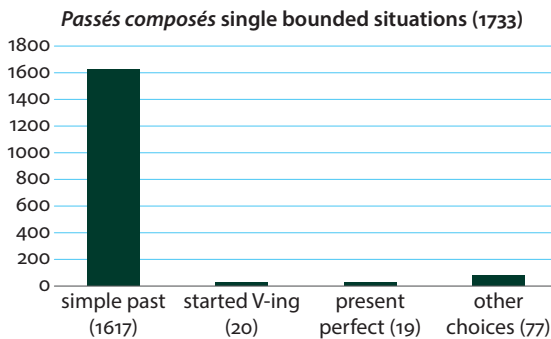


Figure 1. Translations for the single-situation French PC into English

This massive use of the English SP suggests that it is an all-purpose past; this view is reinforced by the comparison with the number of French imperfective pasts (*imparfait*), which totals 1705, only 213 of which are translated as progressive pasts

3. One example of the “*accompli de présent*” of the French PC is:
Et le fait est que la mort de Mme Meursault l’a beaucoup affecté. (Part I, Chapter 1)
And it’s true that Madame Meursault’s death has upset him a great deal.
4. An example of this experiential use of the French PC is:
Il y a des choses dont je n’ai jamais aimé parler. (Part II, Chapter 2)
There are certain things I’ve never liked talking about.

in English. These findings seem to confirm the analysis of de Swart (1998: 365) for whom “the simple past tense in English is aspectually neutral”; it is the only form available for narrating sequences of events. In standard varieties, the English present perfect has not undergone the “aoristic drift” (Squartini & Bertinetto 2000: 417) that the French PC underwent. However, an alternative account proposed by Michaelis (2016) considers the English SP to be aspectually sensitive. One clear illustration of the ambiguous nature of the SP is provided in (1), in which the form ‘ate’ can receive a perfective (a) or an imperfective (b) construal:

- (1) a. FR: *J’ai mangé au restaurant, chez Céleste, comme d’habitude.*
 EN: I ate at Céleste’s restaurant as I always do.
- b. FR: *Nous mangions tous sans parler. Masson buvait souvent du vin.*
 EN: We all ate in silence. Masson drank a lot of wine.

As Michaelis (2016) observes, this distortion of the verb’s *Aktionsart* is akin to well-known coercion effects. Coercion is a term that defines contextual reinterpretations of given eventualities; it arises particularly “when an eventuality description does not meet the input requirements of an aspectual operator, and we get an adjustment, a coerced interpretation of the input, which repairs the mismatch” (de Swart 2000: 7). The viewpoint operators involved in this process are generally progressives and perfects, but Michaelis (2016) contends it also concerns the English SP if it is inserted in a particular discourse environment. The *passé composé*, being perfective and a substitute for the narrative *passé simple* in the novel, sometimes triggers such coercion when it is combined with a stative eventuality such as *penser* ‘think’. For example, in (2a), *j’ai pensé*, ‘I thought’ denotes a bounded albeit durative event in the novel (witness the activity verb *réfléchir* ‘reflect’ that follows it immediately). However, in (2b), the PC with the same verb *penser* ‘think’ which follows one of the few *passés simples* in the text (*passèrent* ‘passed by’), induces a coerced interpretation of the ingressive type (*J’ai pensé* here means ‘it occurred to me’):

- (2) a. FR: *J’ai pensé aux collègues du bureau.(...) J’ai encore réfléchi un peu à ces choses...*
 EN: I thought about my colleagues at work. (...) I thought about that a little more...
- b. FR: *Un peu plus tard passèrent les jeunes gens du faubourg ... J’ai pensé qu’ils allaient aux cinémas du centre.*
 EN: A little while later, some local young men passed by... I thought they were probably going to see a movie in town.

In the novel, 20 occurrences of *j'ai pensé* are translated as 'I thought', exhibiting the same ambiguity as in French, a form that is either a durative past situation of thinking (2a) or a new thinking event (2b), which confirms Michaelis's observation that the English past tense is aspectually sensitive.⁵

In spite of the overall aspectual similarity between the PC and the SP, careful examination of the data shows that the French PC comes out as a near-universal boulder, being available for all kinds of single-situation perfectivity such as telic (*j'ai bu* 'I drank', *j'ai mangé* 'I ate'), change-of-state (*il est devenu* 'he became', *j'ai compris* 'I understood', *j'ai pensé* 'I thought'), as well as limitative (*j'ai marché* 'I walked', *j'ai souri* 'I smiled'). This effect is strengthened by the systematic opposition with the unbounded *imparfait*. In contrast, although aspectually ambiguous, as described above, the English SP is less of a boulder than the French PC, especially because it has no imperfect tense to be contrasted with (cf. the figures for the progressive past vs. the *imparfait*). Therefore, in line with the research on *Aktionsart* marking (Verkuyl 1989; Tenny 1994; Krifka 1998; Borer 2005), we observe that the corpus has many examples of particular constructions or devices that help ensure the appropriate (single-situation perfective) interpretation with the SP: changing the lexical verb (3, 4), the 'started + V-ing' construction (5), special complement constructions (6):

- (3) a. FR: *J'ai pensé que je l'avais déjà dit à mon patron.*
EN: I **remembered** I'd already said that to my boss.
- b. FR: *J'ai pensé que c'était dimanche.*
EN: I **realized** it was Sunday.
- (4) a. FR: *Puis j'ai encore dormi.*
EN: Then I **went back to sleep**
- b. FR: *Il a touché du doigt un bouton sur la table.*
EN: He **pressed** a buzzer on his desk.
- c. FR: *Puis nous avons marché et traversé la ville.*
EN: Then we **went for a walk** in the city.
- (5) a. FR: *Elle a consulté fiévreusement la carte.*
EN: She **started** busily **studying** the menu.
- b. FR: *Il a frappé plus fort et au bout d'un moment, la femme a pleuré et Raymond a ouvert.*
EN: He **banged** on it **harder** and, after a moment, the woman **started crying** and Raymond opened the door.

5. But note that another, older, translation of the novel (Stuart Gilbert's 1946 translation) tends to choose a different verb for durative vs. eventive *penser*. (2a) is 'I **thought of** the other fellows', (2b) is 'I **guessed** they were going'. Whether this shows an evolution of the English language or a different translation decision is not within the scope of this paper to decide.

- c. FR: *Il a sorti du boudin, il l'a fait cuire à la poêle, et il a installé des verres (...).*
 EN: He took out the black pudding and **started cooking** it in a frying pan; he put out the glasses (...).
- (6) a. FR: *J'ai fait ma cuisine et j'ai mangé debout.*
 EN: I prepared my meal and ate **it** standing up.
- b. FR: *Il l'a ôté quand la bière a passé.*
 EN: He took it off as the coffin came **through the door**.
- c. FR: *Brusquement, il s'est levé, a marché à grands pas vers une extrémité de son bureau.*
 EN: Suddenly, he got up, strode **over** to the other end of his office.

Sentences (3) and (4), with the atelic verbs *penser* 'think', *dormir* 'sleep', *toucher* 'touch', *marcher* 'walk' are examples of limited perfectivity: in certain contexts the English verbs are sometimes so to speak too imperfective to be used as such, hence the need to resort to other verbs or constructions to obtain a perfective interpretation. The same constraint, that is to ensure a perfective reading (which here would be ingressive), leads the translator to choose a specific construction ('start V-ing') to focus on the onset for aspectually ambiguous event classes, to avoid either a stative (imperfective) construal ('she studied the menu busily') or a final (eventive) construal ('she cooked it'). Again, this choice is certainly dictated by the fact that the English SP is aspectually ambiguous and can always receive an imperfective construal. In other words, 'she cooked it' may be truth-conditionally equivalent to 'she was cooking it',⁶ which is not the case in French, where *il l'a fait cuire* can only be eventive *vs.* the imperfect *il le faisait cuire*, which is clearly imperfective.

Finally, another frequent strategy consists in adding a bounding NP as in (6a), or in adding a PP satellite to the verb (6b, c) to disambiguate between a stative or eventive interpretation, whereas the French PC is enough in and of itself to ensure the latter. It must be noted that for this last example, which features a motion verb, English uses a complex accomplishment ('strode over to') whereas French, being a verb-framed language (Talmy 2000), relies on the tense-aspect operator (the PC) to make the eventuality bounded (but not lexically telic).

To sum up the discussion on French and English and anticipate the examination of the Russian and Hungarian data, we are led to conclude on the near universal perfectiveness of the French PC: in its aoristic use (that is, equivalent of the *passé simple*), it coerces the situation to be a singularized one. In contrast, the English

6. The following example illustrates the possible imperfective construal of a verb like *cooked*:
Belinda found her way to the kitchen where the housekeeper, Mrs Porter, was creating deliciously savoury aromas as she cooked a three-course evening meal that was carefully balanced with Faye's special needs in mind. (BNC)

SP has been found to be more aspectually ambivalent, hence the observation that several constructions may be used to ensure single-situation (eventive) interpretations. Croft's (2016)⁷ conclusion that "in English, lexical aspectual type is actually a construal of a verb/event in a particular tense-aspect [or other] construction", is confirmed by the analysis presented here. This aspectual ambiguity never happens in Russian, where prefixation entails telicization of the verb's imperfective root: Russian perfective prefixation is strongly coercive. It is the topic of the following section.

3. Russian vs. Hungarian

3.1 The predominance of perfective verbs in Russian

The near universal 'perfectiveness' of the French PC (and the English SP, to a lesser extent) contrasts sharply with Russian, where prefixation of a bare verb⁸ like *kurit* 'smoke' is obligatory to ensure a telic *cum* quantised interpretation (7), but impossible if the reference is bounded-cumulative (8), and both sentences have a single-situation interpretation:

- (7) FR: *J'ai fumé deux cigarettes, je suis rentré.*
 RU: *Vy-kuriv⁹ dve sigarety, ja vernulsja v komnatu.*
 PREF-smoke-PFV-GER two cigarettes I return-PFV-PST to room
 'I smoked two more cigarettes, went inside...'
- (8) FR: *J'ai fumé ensuite des cigarettes, toujours couché, jusqu'à midi.*
 RU: *Valjalsja v posteli do 12, kuril sigarety.*
 lie-IPFV-PST in bed till 12 smoke-IPFV-PST cigarettes
 'Then I smoked in bed until noon.'

The overall results for Russian are as follows: out of the 1733 occurrences of single situation PCs, 81.5% feature a perfective (PFV) verb and 9% an imperfective (IPFV)

7. This quotation by Croft is taken from his plenary talk at the Conference 'Tense, aspect, modality, evidentiality: comparative, cognitive, theoretical, applied perspectives' – held at Université Paris Diderot, November 16–18, 2016.

8. Most bare stems in Russian are imperfective; there are exceptions, however, especially among transfer, ballistic and some change-of-position verbs, where the bare (prefixless) stem is perfective: *dat* 'give'; *kupit* 'buy'; *brosit* 'throw'; *xvatit* 'seize'; *leč* 'lie down'; *sest* 'sit down'; *past* 'fall'.

9. For ease of reading, I will insert a hyphen between the prefix and the stem in the 'prefix + stem' construction; in normal usage, the two are written as one word.

verb. The rest of the occurrences (115) correspond to other translation choices. Among these other choices, a sizable number features non-finite forms (gerunds and participles), which also play a non-trivial role in phenomena of sequence *vs.* overlap interpretations in Russian (Gehrke 2003), and somewhat blur the nice picture (*viz.*, PFV = foregrounding and sequencing *vs.* IPFV = backgrounding and overlap) developed in this chapter. The following two sentences illustrate:

(9) FR: *À ce moment, il s'est tourné vers moi et m'a désigné du doigt...*

RU: *Tut on povernulsja ko mne i, ukazyvaja na*
 there he turn-PF-PST to me and point-IPFV-GER to me
menja pal'cem, ...
 finger-INSTR

'He turned towards me and pointed at me...'

(10) FR: *C'est à ce moment que je me suis aperçu qu'ils étaient tous assis en face de moi à dodeliner de la tête*

RU: *Ja obratil vnimanie, čto kivali oni, usevšis'*
 I turn-PFV-PST attention that nod-IPFV-PST they PREF-SIT-PFV-GER
naprotiv menja.
 opposite me

'Just then I noticed they were all nodding their heads, sitting opposite me'

In sentence (9), an IPFV gerund (*ukazyvaja* 'pointing') is used to ensure a sequential interpretation ('he turned to me and pointed') contra Borik (2002) who believes that only the PFV can have that function, as discussed in (Gehrke 2003: 6). Conversely, a PFV gerund (*usevšis* 'sitting, having sat') can indicate backgrounding (overlap), as in sentence (10), an observation made by Gehrke (2003) for Czech and Russian. This phenomenon is outside the scope of the chapter, which focuses on finite verb forms, but should be taken into account when studying narrative progression.

To come back to the overall results, the past PFV verb is either the PFV partner of a real aspectual pair (i.e., for telic-terminative verbs),¹⁰ or an *Aktionsart* ('mode of action', in the Russian terminology) PFV verb, with such prefixes as delimitative *po-*, ingressive – *za*,¹¹ etc.

10. Examples are (in the order: PFV/IPFV): *sdelat'/delat'* 'do'; *povtorit'/povtorjat'* 'repeat'; *ulybnut'sja/ulybat'sja* 'smile'. These aspect-changing prefixes have been referred to in the literature of aspectology as 'empty' or 'purely perfectivizing' prefixes (e.g. Tixonov 1998; Avilova 1976).

11. Examples: *govorit'* (IPFV), *zagovorit'*, *pogovorit'* (PFV), 'speak', 'start to speak', 'speak for a while'. The prefixes that only change the *Aktionsart* of the main verb are called 'superlexical' or 'external' (in Babko-Malaya 1999; Svenonius 2004; Ramchand 2004; Di Sciullo & Slabakova 2005)

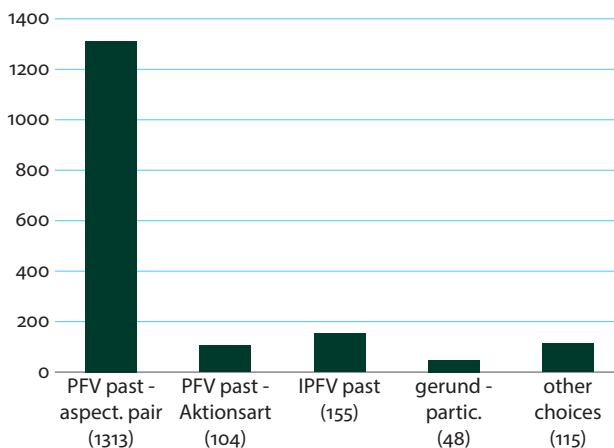


Figure 2. Translations for the single-situation French PC into Russian

The overwhelming number of PFV pasts as translations for single-situation French PC forms, testifies to a typological feature of Russian, which has a binary obligatory grammaticized opposition (Maslov 1958; Bondarko 1971; Zaliznjak & Šmelev 1997). These results confirm the nature of the Russian PFV, which is of the “complete” type with a strong bias for marking single bounded events (Plungian 2012). The expression of limitativeness with the help of inseparable spatial modifiers of the verb (so-called preverbs) is a widespread phenomenon (Schooneveld 1978; Borer 2005; Janda 2013). However, there is a tendency for the number of productive preverbs to decrease, and consequently, a rise of what Janda (2013) calls “classificatory perfectivization”. In the corpus, it is indeed the case that many ‘prefix + stem’ verbs (11a) are simply non-compositional, because the prefixless (IPFV) root disappeared from the language or has become obsolete as in (11b):

- (11) a. *umeret* ‘die’, *polučit* ‘receive’, *skazat* ‘say’, *prinjat* ‘take’, *ob’jasnit* ‘explain’, *pomoč* ‘help’, etc.
 b. *?meret*, **lučit*, **kazat*, **jat*, **jasnit*, *#moč*

This fact demonstrates to the strongly grammatical nature of Russian aspect, which is not instantiated in Hungarian, where the detachable prefix is still lexical, as we will see in Section 3.

3.2 Sensitivity to the lexical nature of the predicate

For the expression of single situations, PFV verbs are dominant across all *Aktionsart* classes, as we said in the preceding section. But it is also true that the proportions of PFV verbs over IPFV verbs are directly tied to the *Aktionsart* nature of the lexical predicate.



Figure 3. Sensitivity of Russian PFV and IPFV verbs to the *Aktionsart* class of the predicate

Figure 3 shows that accomplishments and achievements account for 77% of the occurrences of all PFV verbs, while the 33% remaining occurrences of PFV verbs feature state and activity verbs, which is non negligible. The sentence in (12) is a good illustration of this observation:

(12) FR: *Raymond a hésité, m'a regardé et a tiré sur sa cigarette.*

RU: *Rajmon za-mjalsja, po-gljadet na menja i
Raymon PREF-hesitate-PFV-PST PREF-look-PFV-PST at me and
za-tjanulsja sigaretoj.
PREF-draw-PFV-PST cigarette-INSTR*

'Raymond hesitated, glanced over at me and took another drag of his cigarette.'

The atelic *po-* verb (*po-gljadet* 'look, glance') fulfils the same function of narrative advancement as the telic events *zamjalsja* 'hesitated' and *zatjanulsja* 'take a drag'. The prefix *po-* has an important grammatical function in the aspectual system of Russian as it allows "the extension of the aspect opposition to atelic activity predicates ... [it] allow[s] activity predicates to be sequenced in time on a par with telic predicates" (Dickey 2006: 16). In fact the 33% of Russian PFV states and activities found as translations of single situation French PCs are coerced constructions, what the Russian tradition calls mode of action perfectivity: added to an IPFV root, the 'superlexical' prefix (*po-*, *za-*, *u-*, etc.) adds a bound to an otherwise atelic verb as shown in the following examples:

- (13) FR: *Puis je suis rentré chez moi, j'ai dormi un peu parce que j'avais trop bu de vin et, en me réveillant, j'ai eu envie de fumer.*

RU: *Potom zabežal domoj, po-spal*
 then I.run-PFV-PST home po-sleep-PVF-PST a.little
nemnogo, potomu čto xvatil lišnij stakančik vina.
 because that I.take-PFV-PST extra glass of.wine when
Kogda prosnulsja, očen' za-xotelos' kurit'.
 I.wake.up-PFV-PST a.lot za-want-PFV-PST smoke-INF

'Then I went home and slept a little because I'd had too much wine with lunch, and when I woke up, I felt like having a cigarette.'

- (14) FR: *Un peu plus tard, pour faire quelque chose, j'ai pris un vieux journal et je l'ai lu.*

RU: *Poxodiv, ja vzjal staruju gazetu,*
 walk-PFV-GER I take-PFV-PST old newspaper
po-čital eë.
 I.po-read-PFV-PST it

'A little while later, because I had nothing else to do, I picked up an old newspaper and read it.'

- (15) FR: *Quand elle a ri, j'ai eu encore envie d'elle.*

RU: *Kogda Mari za-smejalas', ja opjat' eë za-xotel.*
 When Mari za-laugh-PFV-PST I again her za-want-PFV-PST
 'When she laughed, I wanted her again.'

- (16) FR: *Dès qu'il m'a vu, il s'est soulevé un peu et a mis la main dans sa poche.*

RU: *Kak tol'ko arab u-videl menja, on pripodnjalsja i*
 as soon as Arab u-see-PFV-PST me he raised.himself and
sunul ruku v karman.
 put hand in pocket

'As soon as he saw me, he raised himself up a bit and put his hand in his pocket.'

In the preceding examples, the PFV verbs are translations of PC verbs in French: *pospal* 'slept', *zaxotelos'* 'wanted', *počital* 'read', *zasmejalas'* 'laughed', *uvidel* 'saw, caught sight of', are all modes of action. The situation is simply bounded, the PFV of these otherwise atelic verbs is required by the context of narrative sequencing ('X did Y, then did Z...') and conveys meanings such as ingression (as in examples 15 and 16), limited duration or lack of specific result (as in examples 13 and 14). They all contribute a new development in the narrative. But more strikingly, it must be noted that the number of IPFV verbs for the same state and activity verbs is also very frequent, even in the narration: Figure 3 shows that 11% of single-situation PCs for state and activity verbs correspond to IPFV verbs. In other words, atelics

like *j'ai dormi* 'I slept', *j'ai eu envie* 'I wanted', *j'ai lu* 'I read', *j'ai vu* 'I saw' can be translated by a PFV verb (*Ja pospal, mne zaxotelos', ja počital, ja uvidel*) as well as by the corresponding prefixless IPFV verb, *ja spal, mne xotelos', ja čital, ja videl*. In Examples (17) to (21), which feature the same verb in the French PC as in Examples (13) to (16), the Russian translation has IPFV verbs:

- (17) FR: *J'ai dormi pendant presque tout le trajet.*
 RU: *Ja spal počti vsju dorogu.*
 I sleep-IPFV-PST almost all-ACC road-ACC
 'I slept for nearly the whole journey.'
- (18) FR: *Elle a eu un petit recul, mais n'a fait aucune remarque. J'ai eu envie de lui dire que ce n'était pas ma faute, mais je me suis arrêté (...).*
 RU: *Ona čut'-čut' otrjanela, no ničego ne skazala.*
 she just flinch-PFV-PST but nothing not she.say-PFV-PST
Mne xotelos' skazat': Ja tut ne vinovat', odnako ja
 1SG-DAT want-IPFV-PST say I here am.not guilty however I
promolčal...
 keep.quiet-PFV-PST
 'She flinched a little but didn't say anything. I wanted to tell her that it wasn't my fault, but I stopped myself (...).'
- (19) FR: – *"Moi aussi, a-t-il dit, je me suis penché sur cette âme, mais (...) j'ai trouvé quelque chose et je puis dire que j'y ai lu à livre ouvert"*.
 RU: – *Ja tože zagljanul v ètu dušu, – skazal on, – (...) no ja*
 I too look-PFV-PST into this soul said he but I
mnogoe našël v nej i mogu skazat', čto ja čital
 a.lot find-PFV-PST in it and can say that I read-IPFV-PST
v nej kak v raskrytoj knige.
 in it like in open book
 'I too,' he said, 'have looked into this man's soul, but (...) I did find a soul, and I can truthfully say that to me it was an open book.'
- (20) FR: *Il est allé vers les journalistes, a serré des mains. Ils ont plaisanté, ri ...*
 RU: *On napravilsja k žurnalistam, obmenjalsja s nimi*
 he go.over-PFV-PST to journalists exchange-PFV-PST with them
rukopožatijem. Oni perekidyvalis' šutkami, smejalis'...
 handshaking they throw-IPFV-PST jokes laugh-IPFV-PST
 'He went over to the journalists and shook their hands. They exchanged pleasantries, laughing...'

- (21) FR: *Au large, nous avons fait la planche (...). Nous avons vu que Masson regagnait la plage pour s'étendre au soleil.*

RU: *My zaplyli daleko i legli na spinu*
 we swim.out-PFV-PST far and lie-PFV-PST on back
 (...) *My videli, čto Masson vybralsja na bereg i*
 we see-IPFV-PST that Masson go.onto-PFV-PST on beach and
rastjanulsja tam, grejas' na solnce.
 stretch.out-PFV-PST there basking in sun

'Once we were out in the open, we floated on our backs (...). We saw that Masson had gone back on to the beach to stretch out in the sun.'

This is one of the thorniest issues when dealing with the past tense in Russian: an IPFV past tense verb is polyfunctional in that it allows many different readings such as habitual, progressive, perfect, and the so-called Imperfective General Factual (IGF; Forsyth 1970; Comrie 1976; Glovinskaja 1982; Leinonen 1982; Padučeva 1996; Grønn 2004). An elaborate discussion of the factual imperfective is found in Grønn (2004); it denotes a complete past event and concerns telic verbs, and so it is a case of aspectual competition with the PFV. Grønn (2004: 223) distinguishes two types: on the one hand, the existential IPFV, with a "big and floating' past assertion time", which contrasts with the "definite, more restricted and narrow assertion time" (ibid.) of the PFV, and on the other hand, the presuppositional IPFV, linked by anaphora to a previous event in the discourse. We find both types in the data. What makes the factual IPFV a frequent choice in *L'Etranger* is precisely the lack of temporal definiteness permitted by the use of the French PC, as often noted. As soon as there is a gap or breach in the narrative progression (Grønn 2004: 142) and the situation becomes temporally indefinite (or non-specific), the factual IPFV occurs.

Examples (17) to (21) illustrate this use, which can be explained by different factors. In (17), it is the durative time adverbial (*vsju dorogu*, 'the whole journey') and the insertion of the atelic event of sleeping outside of the main storyline that favors the IPFV. In (18), the speaker presents an isolated situation, not contingent upon a previous situation, and an action that has actually aborted. The sentence in (19) appears in direct speech, the focus is on the activity of the prosecutor (figuratively reading in Meursault's soul) at some non-specific time in the past, with the result or target state of the event irrelevant: only the fact of reading in the soul matters. In (20), the pleasantries and the laughing are not considered as a sequence of chronologically ordered episodes, they do not advance the narrative, but describe the potentially iterative activity of the characters at a given moment in the narrative. This example may also correspond to Asher and Lascarides' (2003) SDRS of *Parallel*: the consecutive occurrences of factual IPFV "are not temporally related to each other [as in *Narration*], but they are still coordinated". Finally, in (21), the IPFV

verb *videl* precludes the ‘catch sight of’ interpretation, not relevant here: the vision of the characters is backgrounded to the actual event of interest (Masson having gone back to the beach), and here again we note the gap in the temporal structure, in contrast with the sentence in (16) where the PFV verb *uvidel*, ‘saw’, maintains a strict temporal progression between the events.

The other type of factual IPFV, the presuppositional (Grønn 2004), is illustrated in the sentence in (23). Note that the same truth-conditional situation is expressed by the PFV verb in the sentence in (22); this is a typical case of aspectual competition, and concerns telic-terminative (accomplishment) verbs:

(22) FR: *Alors, j’ai tiré encore quatre fois sur un corps inerte où les balles s’enfonçaient sans qu’il y parût.*

RU: *Togda ja vy-strelil eščë četyre raza v nepodvižnoe telo, v*
 then I fire-PFV-PST again four times into lifeless body into
kotoroe puli vonzalis’ nezametno.
 which bullets sink-IPFV-PST without.a.trace

‘Then I fired for more times into the lifeless body, where the bullets sank without leaving a trace.’

(23) FR: “*Pourquoi, pourquoi avez-vous tiré sur un corps à terre?*”

RU: *Počemu? Počemu vy streljali v rasprostërtoe na zemle,*
 why why you fire-IPFV-PST into lying on ground
nepodvižnoe telo?
 lifeless body

‘Why, why did you fire at a man who was already dead?’

The PFV in (22) pertains to the “assertoric level” of discourse, as opposed to the “presuppositional level” for (23). For Grønn (2004: 224), the assertoric level has the function of introducing “new events” in the storyline: it is the case in (22), where the narrator has to give a complete restatement of the fact (‘the shooting of the victim’), which goes with accentuation of the PFV verb *vystrelil* ‘shot’. However, the sentence in (23) has a different discourse function: the event of shooting is already given in the context – that is the presuppositional level – the IPFV verb is de-accented, and the focus is on the circumstances that surround the event (here, the *why* bears sentential stress). That is a discursive-pragmatic distinction that the French PC alone does not encode.

Therefore, of utmost importance in Russian for the choice of a PFV or IPFV verb in a narrative discourse mode is whether the predicate constitutes a (chrono) logical development in the narrative or not: a single situation may still be foregrounded but not sequentially tied to a preceding episode, and the factual IPFV can occur. The Russian IPFV for atelics can thus have two functions: providing a

backgrounded stative episode (like the French *imparfait*) or a foregrounded one but not explicitly linked to the development of the narrative, as shown by examples (17) through (21), which highlight the importance of information structure for aspectual choice in Russian.

3.3 Hungarian

Finally, we look at the Hungarian data and try to address a controversial question in the literature: does Hungarian have Slavic-like aspect (Kiefer 1994; Csirmaz 2006) or not (Kiss 2006)? Observation of the corpus for the expression of single situations provides some answers. First, it must be noted that Hungarian has morphological features which are very similar to those of Russian: each verb has one past tense *-t(t)-* morpheme (24), verbs come in two guises, bare (prefixless) or prefixed¹² (25), with an important difference, though: the prefix is separable, contrary to Russian, with two configurations: ‘prefix + stem’ vs. ‘stem + prefix’ (26).¹³

(24) *néz* ‘look’ → *néz-t-em* ‘I looked’, *néz-ett* ‘he looked’

(25) *megy* ‘go’ → *el-megy* ‘go away’, *fel-megy* ‘go up’
iszik, ‘drink’ → *meg-iszik* ‘drink up’, *ki-iszik* ‘drink out’

(26) FR: *Aujourd’hui, maman est morte. Ou peut-être hier, je ne sais pas. J’ai reçu un télégramme de l’asile: “Mère décédée...”*

HU: *Ma halt meg anyám. Vagy talán tegnap,*
 today die-PST PREF my.mother or maybe yesterday,
nem is tudom pontosan. A menhelyből
 not at.all I.know exactly the old.people’s.home-from
sürgönyözték: “Anyja meg-halt”.
 they.cabled your.mother PREF-die-PST

‘My mother died today. Or maybe yesterday, I don’t know. I received a telegram from the old people’s home: “Mother deceased”’

12. The terminology is fluctuating; in traditional Hungarian terminology, the particle-like element that attaches to the verb is called a ‘verbal prefix’, but some authors (e.g. Kiss 2006; Surányi 2009) prefer to use the term ‘verbal particle’ or ‘preverb’. As nothing hinges on the choice of one term over the other, and to keep the parallel with Slavic prefixes, I will call this element ‘(verbal) prefix’.

13. Just as I did for the Russian data, for ease of reading I will insert a hyphen between the prefix and the stem in the ‘prefix + stem’ construction; in normal usage, the two are written as one word but are separable, contrary to Russian.

3.3.1 Sensitivity to Aktionsart

The most striking finding for Hungarian is that, contrary to Russian, we find an almost equal distribution between bare and prefixed verbs in the expression of single situations: 48.2% of French PC single-situation forms are translated by bare verbs, and 45.2% by prefixed verbs.

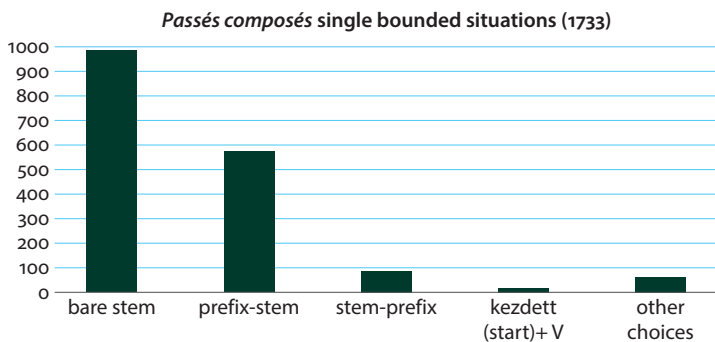


Figure 4. Translations for the single-situation French PC into Hungarian

Even more marked than for Russian, we note a strong sensitivity between the *Aktionsart* nature of the predicate and the bare *vs.* prefixed form of the verb: the correlation between telic situations (accomplishments and achievements) and prefixation is indeed important (82.8%), but there is a sizable number of prefixed atelics as well (17%). Unlike Russian, bare verbs are well represented for each class, indeed predominant for atelics (58.5%), but non negligible for telics (41.5%).

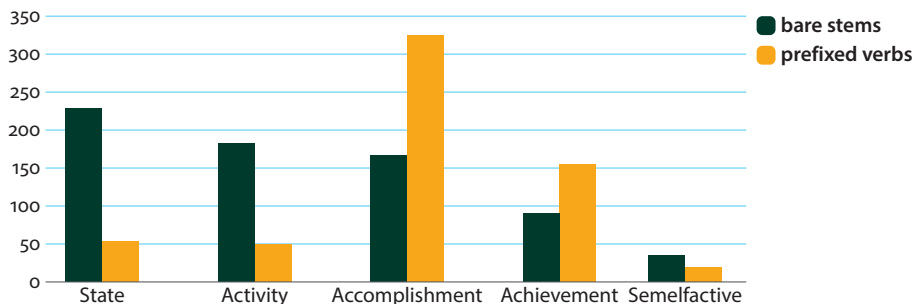


Figure 5. Sensitivity of Hungarian bare and prefixed verbs to the *Aktionsart* class of the predicate

An illustration of this complex behavior of verbs in Hungarian and Russian with respect to prefixation is provided in (27).

(27) FR: *Il a sorti du boudin, il l'a fait cuire à la poêle, et il a installé des verres, des assiettes, des couverts et deux bouteilles de vin.*

HU: *Elő-kereste a hurkát, meg-sütötte a serpenyőben, s poharakat, PREF-get-PST the pudding PREF-fry-PST in pan and glasses tányérokot evőeszközöket hozott, azonkívül két plates cutlery bring-PST what's.more two bottle üveg bort is. wine too*

RU: *On dostal krovjanuju kolbasu i pod-žaril eë na skovorode, he get-PFV-PST pudding and fry-PFV-PST it on pan pri- nës stakany, tarelki, pribory, (...) postavil dve butylki bring-PFV-PST glasses plates cutlery put-PFV-PST two bottles vina. of.wine*

'He took out the black pudding and started cooking it in a frying pan; he put out the glasses, plates, cutlery and two bottles of wine.'

All three verbs in (27) are telic accomplishments in a narrative sequencing: Russian has three PFV prefixed verb,¹⁴ with no possibility of alternating with the corresponding prefixless verbs, because that would automatically make the situations imperfective (progressive). Hungarian, however, has only two prefixed verbs (*elő-kereste* 'he got out', *meg-sütötte* 'he cooked') followed by definite objects, which also contribute to measuring out the event and make it telic. But the last verb in the series is a bare stem with an unambiguously telic interpretation (*hozott* 'he brought'), and the immediate pre-verbal position is occupied by bare nominals ('glasses, plates, cutlery brought').

There has been an abundant literature on the comparable syntax of bare nominals and their various aspectual interpretations as in Kiefer (1994), as incorporated objects in Farkas & de Swart (2003) and verbal prefixes, with no definite conclusion on the semantics with respect to aspect marking. For example, Kiss (2002: 69), drawing on the parallel distribution of bare nominals, verbal prefixes and resultative phrases, writes that these "represent the same functional class in syntax – even if the precise semantic content of their various types (...) is not fully understood at present". But (Kiss 2006) tones down this initial claim; she merely observes that different classes of verbs with (preverbal) bare nominals behave differently with respect to aspect: a sentence like *Péter fát vágott* ('Peter wood cut') "merely denote[s]

14. The first one, *dostal* 'took out' is etymologically prefixed (*do + stal*) but in its current use it is non-compositional; the IPFV verb is obtained by suffixation (*dostavat*).

[a] process, which can be interpreted either imperfectly ('he was cutting the wood') or perfectly ('he cut the wood')" (Kiss 2006: 22–23). This corresponds to my observations about the corpus: the aspectual contribution of bare (incorporated) nominals must be relativized to the lexical class of the verb. Therefore, in what follows, I will take seriously the division of verbs into different lexical classes, which behave differently with respect to aspect but will make no claim about the syntax of bare nominals and prefixes, as this requires further research.

3.3.2 *Different classes of verbs*

This importance of a verb's lexical class in disentangling the aspectual interpretation of sentences in the corpus confirms what the literature has brought to the fore fairly recently (Bende-Farkas 2002; Kiss 2006; Piñon 2006; Kardos 2011), that for certain lexical classes of verbs (mainly, appearance, creation, and consumption verbs), telicity or bounding is obtained **without** prefixation, as shown in (28), (29) and (30):

- (28) FR: *À ce moment, un camion est arrivé.*
 HU: *Épp abban a pillanatban egy teherautó érkezett...*
 just in the instant a truck arrive-PST
 RU: *Kak raz tut pod-'exal gruzovik.*
 just then here PREF-arrive-PFV-PST truck
 'Just then, a truck arrived'.
- (29) FR: *J'ai encore pris du café au lait qui était très bon.*
 HU: *Megint ittam tejeskávét, s megint igen jólesett.*
 again I.drink-PST milk.coffee and again very did.good
 RU: *Potom ja opjat' vy-pil očen' vkusnogo kofe*
 then I again PREF-drink-PFV-PST very good coffee
s molokom.
 with milk
 'I had another coffee; it was very good'.
- (30) FR: *Je t'ai acheté un ensemble ce mois-ci.*
 HU: *Nemrég kaptál tőlem kosztümöt.*
 not.long.ago 2SG.receive-PST from.me outfit
 RU: *Ja tebe kupil v ètom mesjace kostjumčik.*
 I you.DAT buy-PFV-PST in this month outfit
 'I bought you a new outfit this month'.

The prefixless verbs *érkezik*, 'arrive', *iszik*, 'drink', and *kap*, 'get, receive', belong to a class of verbs that Bende-Farkas (2002: 8) calls "definiteness effect (DE) verbs": they express the coming into existence or into being (or the opposite, the going out of existence) of an entity, and are found among very different lexical semantic

classes of verbs: appearance, creation, consumption, destruction verbs. They behave like DE verbs only if they co-occur with an *indefinite* object as in (29, 30) or an unaccusative subject as in (28); due to their lexical semantics, which is basically existential, this construction – ‘unprefixed stem + indefinite NP’ – correlates with “predicate novelty” (ibid.: 224). In other words, just as (28) means that a truck came into view by arriving, the sentence in (29) means that coffee went out of existence because of the subject’s referent drinking it, and the sentence in (30) means that a costume came into existence as a result of an event of receiving. This particular morphosyntactic behavior implies that in all three cases, we are dealing with new events. But it is the construction as a whole (*i.e.* ‘unprefixed stem + indefinite NP’) that creates this effect, not the verb: in an imperfective context, the same prefixless verb (*ivott*, ‘he.drank’)¹⁵ would be used, as in (31), which repeats Example (1). Again, Russian has to use an IPFV verb (*pil* ‘drank’) in such a context:

- (31) FR: *Nous mangions tous sans parler. Masson buvait souvent du vin.*
 HU: *Mindannyian szó nélkül faltunk. Masson igen sok*
all word without eat-PST-1PL Masson indeed a.lot.of
bort ivott.
wine drink-PST-3SG
 RU: *Vse eli molča. Masson pil mnogo vina.*
all eat-IPFV-PST silently Masson drink-IPFV-PST a.lot.of wine
 ‘We all ate in silence. Masson drank a lot of wine’.

Interestingly, Hungarian behaves more like English in this sentence: the form ‘drank’ can be perfective or imperfective, just like the bare verb *ittam/ivott* in Hungarian. This stands in sharp contrast with Russian, where all three translations of these verbs feature PFV verbs, two of them being prefixed (28, 29), the other (30) non-prefixed but PFV nonetheless.

In contrast, the prefix added to these same verbs, followed by a definite NP object, cancels the existential flavor of the sentences and creates what Bende-Farkas (2002) and Piñon (2006) call a “specificity effect”; “prefixed verbs (...) confirm previous knowledge about discourse entities” (Bende-Farkas 2002: 224; 86), as in sentences (32), (33) and (34):

15. *Ivott*, ‘he.drank’ is the third-person past tense form of *ittam* ‘I.drank’, which appeared in (29).

- (32) FR: *Mon avocat est arrivé, en robe, entouré de beaucoup d'autres confrères.*
 HU: *Meg-érkezett az ügyvéd is, talárossan, egy csomó*
 PREF-arrive-PST the lawyer too in.robe a plenty.of
kartárs között.
 colleagues among
 'My lawyer arrived wearing his robe and surrounded by several of his colleagues.'
- (33) FR: *Il a cacheté la lettre et nous avons fini le vin.*
 HU: *A levele-t leragasztotta, s meg-ittuk*
 the letter-ACC he.glue-PST and PREF-drink-PST-1PL
a maradék bort.
 the rest of the wine
 'He put the letter in the envelope and we finished the wine.'
- (34) FR: *Je lui ai dit qu'il me semblait que maintenant elle était punie (...).*
 HU: *Mondtam neki, hogy a lány most már meg-kapta*
 tell-PST-1SG him-DAT that the girl now already PREF-receive-PST
*a büntetését*¹⁶ (...).
 the punishment-POSS3
 'I told him that it seemed to me that she'd been punished (...)'

With those verbs, the prefixed version (often the prefix *meg-*) has presuppositional meaning (Kardos 2011), which implies that their internal arguments must be familiar from the discourse context and they are presupposed. For our purposes, the important observation, confirmed by the high number of accomplishments and achievements for the prefixless verbs (see Figure 5), is that for this class of verbs, the prefix does **not** create telicity or bounding, but has to do with discourse parameters like the quantity of information available on the event. Hungarian is sensitive to topic-focus structure, it is a "discourse configurational language" (Kiss 1995: 5): any investigation into its aspectual system depends very much on uncovering the principles that regulate focus, a goal beyond the scope of this chapter. What the corpus also suggests is that this behavior of appearance, creation, consumption verbs with respect to the presence *vs.* absence of the prefix, may also be found in other verb classes, such as change of location sentences. The corpus contains several examples of a prefixless verb preceded by a preverbal – that is in focus position – directional PP, with a single situation interpretation:

16. The Hungarian translation literally says 'that she received her punishment'.

- (35) FR: *L'on m'a emmené dans la petite pièce (...).*
 HU: *Engem meg abba a kicsike városzobába vezettek (...).*
 1SG-ACC then into the small room lead-PST-3PL
 RU: *A menja u-veli v tu malen'kuju komnatu (...)*
 1SG-ACC PREF-lead-PST-PFV-3PL into that small room
 'I was taken to the small room (...)'
- (36) FR: *Nous avons pris un autobus.*
 HU: *Autóbuszra szálltunk*
 bus-into get.on-PST-1PL
 RU: *My seli v avtobus.*
 we sit-PFV-PST in bus
 'We took a bus.'

The 'directional PP + verb stem' construction in (35) and (36) is in complementary distribution with the 'prefix + stem' construction. They fulfil the same functional role as in sentences (37) and (38) below, in which the prefixes *el-* 'away', and *fel-* 'up', occupy the pre-verbal slot, which is the one for focus, just like the directional PPs in (35) and (36), which ensure event bounding. Again, this never happens in Russian, where in all these cases a PFV or a factual IPFV verb (prefixed in 37, simplex in 38) is the only choice to describe a single episode:

- (37) FR: *C'est alors qu'on m'a emmené.*
 HU: *Akkor aztán el-vezettek.*
 then after PREF-lead-PST-3PL
 RU: *I togda menja u-veli.*
 and then 1SG.ACC PREF-lead-PFV-PST
 'Then I was taken away.'
- (38) FR: *Nous avons pris l'autobus.*
 HU: *Fel-szálltunk az autóbuszra.*
 PREF-get.on-PST-1PL the bus-onto
 RU: *My seli v avtobus.*
 we sit-PFV-PST in bus
 'We took the bus.'

3.3.3 Aspect in HU

These important findings lead us to conclude that in spite of obvious similarities (the presence of telicizing prefixes, the importance of information structure, among others), the Hungarian aspect system is different from the Russian. What the preceding section has established is that prefixation of a bare stem in Hungarian does not automatically entail perfectivity: verbs behave differently according to the

semantic class to which they belong. Moreover, there is no such thing as imperfective suffixation in iterative and progressive contexts in Hungarian: there is no formal difference in the expression of a single perfective episode (39) and a progressive imperfective one (40), something which never happens in Russian, which systematically contrasts a PFV verb (39) with an IPFV verb (40):

(39) FR: *Nous sommes montés et j'allais le quitter (...).*

HU: *Fel-mentünk az emeletre, s már búcsúzni*
 PREF-go-PST-1PL to.upstairs and already say.goodbye
akartam tőle.
 want-PST-1SG from.him

RU: *My podnjalis' vmeste s Rajmonon, i ja uže sobiralsja*
 1PL go up-PFV-PST together with Rajmon and I already got.ready
prostit'sja s nim.
 say.goodbye with him

'We walked upstairs and I was about to say goodbye to him.'

(40) FR: (...) *pendant qu'il gravissait l'escalier de bois, je suis resté...*

HU: (...) *s míg fel-ment a falépcsőn, én lenn maradtam*
 and while PREF-go-PST-3SG on wooden.stairs I down stay-PST-1SG

RU: (...), *i, poka on podnimalsja po derevjanoj lestnice,*
 and while he go.up-IPFV-PST on wooden stairs
ja stojal vnizu.
 I stand-IPFV-PST down

'(...) while he went up the wooden stairs, I stopped ...'

The examples found in the corpus run counter to a claim often made in the literature on Hungarian aspect (Kiss 1995; Kiefer 1994), according to which moving the prefix to a postverbal position ensures a progressive interpretation. In fact, it appears that an imperfective (progressive) construal can be obtained **without** inverting the 'prefix + stem' order, as (40) shows. The conclusion to be drawn from these observations is that Hungarian does not have morphological Slavic-like aspect. It is situation aspect that is grammaticized (Kiss 2006).

4. Conclusion

This corpus-based study has shown that the means of expressing singularized bounded episodes across four different languages are different. The main findings of this empirical study are that in spite of lexical or morphosyntactic similarities between sets of languages (English and French on the one hand, Russian

and Hungarian on the other), each language has developed its own tense-aspect system, which consists in morphemes or constructions interacting with other constraints and constructions of the language under consideration. The French PC comes out as a near universal bounder compared with its unbounded imperfect counterpart. English having no general imperfect form of the Romance type, the SP is aspectually ambiguous, receiving perfective as well as imperfective interpretations. Russian relies on a morphosyntactic construction ('prefix + bare verb') to create single-situation PFV verbs, and that is what has grammaticalized to become Slavic aspect. Hungarian has similar morphosyntactic resources (prefixless and prefixed verbs), but no Slavic-like aspect, which means that the presence or absence of the prefix does not automatically correlate with the expression of bounded *vs.* unbounded situations.

The tense-aspect markers studied in this chapter are mainly type-shift constructions, taking as input a verb with its lexical semantic aspectuality, and yielding tense-aspect interpretations. The difficulty lies in uncovering exactly the nature of that coercion. Considering the object of this study (the expression of single bounded episodes), the closest candidate for a near universal formal marking of single bounded episodes turns out to be Russian, with the PFV verb and its bias towards marking single events. Hungarian appears as the opposite, as single *vs.* multiple events do not receive any formal marking, and the past tense inflection is no coercer, contrary to French or English. However, the present chapter has not made any claim about the tense-aspect systems of the four languages under discussion, as other important ingredients, such as imperfectivity (with its sub-categories of habituality, progressiveness, genericity), or the use of non-finite forms in creating narrative sequencing, have been left out of the discussion, as well as interactions between the syntactico-semantic and the pragmatic dimension of aspectual meaning such as information structure, presupposition, topic-focus, which have only been alluded to. But we hope that the empirical investigation into the expression of single episodes has shed some light on the differences in the encoding of aspectual distinctions.

References

- Asher, Nicholas & Lascarides, Alex. 2003. *Logics of Conversation*. Cambridge: CUP.
- Avilova, Natalia. 1976. *Vid glagola i semantika glagol'nogo slova* (Verb Aspect and the Semantics of the Verbal Word). Moskva: Nauka.
- Babko-Malaya, Olga. 1999. *Zero Morphology: A Study of Aspect, Argument Structure and Case*. PhD dissertation, Rutgers University.
- Bende-Farkas, Agnès. 2002. *Verb Object Dependencies in Hungarian and English: A DRT-based Account*. PhD dissertation, Universität Stuttgart.

- Benveniste, Emile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.
- Bertinetto, Pier Marco. 2001. On a frequent misunderstanding in the temporal-aspectual domain: The ‘perfective-telic’ confusion. In *Semantic Interfaces*, Carlo Cecchetto, Genaro Chierchia & Maria T. Guasti (eds), 177–210. Stanford CA: CSLI.
- Bondarko, Alexander. 1971. *Grammatičeskaja kategorija i kontekst* [Grammatical Category and Context]. Leningrad: Nauka.
- Borer, Hagit. 2005. *The Normal Course of Events*. Oxford: OUP.
- Borik, Olga. 2002. *Aspect and Reference Time*. PhD dissertation, Utrecht University.
- Bybee, Joan & Dahl, Östen. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13: 51–103. doi:10.1075/sl.13.1.03byb
- Bybee, Joan, Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in Languages of the World*. Chicago IL: University of Chicago Press.
- Camus, Albert. 1942. *L’Etranger* (Gallimard, Collection Folio); *The Outsider* (Penguin Books, 2012; translator: Sandra Smith); *Postoronij* (Prometej, 1989; translator: N. Nemčinovoj); *Közöny* (Szépirodalmi, Budapest, 1989; translator: Gyergyai Albert).
- Caudal, Patrick. 2015. The *passé composé* in Old French and Modern French. In *Sentence and Discourse*, Jacqueline Guéron (ed.), 178–205. Oxford: OUP. doi:10.1093/acprof:oso/9780198739418.003.0009
- Comrie, Bernard. 1976. *Aspect*. Cambridge: CUP.
- Croft, William. 2016. The aspectual potential of verbs and lexical semantic typology. Paper presented at the Conference Tense, Aspect, Modality, Evidentiality: Comparative, Cognitive, Theoretical, Applied Perspectives, Université Paris Diderot.
- Csirmaz, Aniko. 2006. Accusative case and aspect. In *Event Structure and the Left Periphery – Studies on Hungarian*, Katalina É. Kiss (ed.), 159–199. Dordrecht: Springer. doi:10.1007/978-1-4020-4755-8_8
- Depraetere, Ilse. 1995. On the necessity of distinguishing between (un)boundedness and (a)telicity. *Linguistics and Philosophy* 18: 1–19. doi:10.1007/BF00984959
- Di Sciullo, Anna-Maria & Slabakova, Roumyana. 2005. Quantification and aspect. In *Perspectives on Aspect*, Angeliek van Hout, Henriëtte de Swart & Henk Verkuyl (eds), 61–80. Dordrecht: Springer. doi:10.1007/1-4020-3232-3_4
- Dickey, Stephen M. 2006. Aspectual pairs, goal orientation and *po*-delimitatives in Russian. *Glossos Issue* 7: 1–37.
- Dik, Simon. 1989. *The Theory of Functional Grammar*, Part I: The Structure of the Clause. Dordrecht: Foris.
- Farkas, Donka & de Swart, Henriëtte. 2003. *The Semantics of Incorporation: From Argument Structure to Discourse Transparency*. Stanford CA: CSLI.
- Forsyth, John. 1970. *A Grammar of Aspect*. Cambridge: CUP.
- Furkó, Péter. 2014. Perspectives on the translation of discourse markers. *Acta Universitatis Sapientiae, Philologica* 6(2): 181–196.
- Gehrke, Berit. 2003. *Aspectual affixes in Russian and Czech*. Ms, Utrecht University.
- Glovinskaja, Marina. 1982. *Semantičeskije tipy vidovyx protivopostavljenij russkogo glagola* (Semantic types of aspectual oppositions in the Russian verb). Nauka: Moscow.
- Gosselin, Laurent. 1996. *Sémantique de la temporalité en français: Un modèle calculatoire et cognitif du temps et de l’aspect*. Louvain-la-Neuve: Duculot.
- Grønn, Atle. 2004. *The Semantics and Pragmatics of the Russian Factual Imperfective*. PhD dissertation, Oslo University.

- Guillemin-Flescher, Jacqueline. 1981. *Syntaxe comparée du français et de l'anglais: Problèmes de traduction*. Paris: Ophrys.
- Hopper, Paul J. 1982. Aspect between discourse and grammar: An introductory essay for the volume. In *Tense-Aspect – Between Semantics and Pragmatics* [Typological Studies in Language 1], Paul J. Hopper (ed.), 3–18. Amsterdam: John Benjamins. doi:10.1075/tsl.1.04hop
- Janda, Laura. 2013. *Why Russian Aspectual Prefixes Aren't Empty: Prefixes as Verb Classifiers*. Bloomington IN: Slavica.
- Kardos, Éva A. 2011. *Toward a Scalar Semantic Analysis of Telicity in Hungarian*. PhD dissertation, University of Debrecen.
- Kiefer, Ferenc. 1994. Some peculiarities of the aspectual system in Hungarian. In *Tense, Aspect and Action – Empirical and Theoretical Contributions to Language Typology* Carl Bache & Carl Basbøll (eds), 185–205. Berlin: Mouton de Gruyter. doi:10.1515/9783110883077.185
- Kiss, Katalina É. 1995. *Discourse Configurational Languages* [Oxford Studies in Comparative Syntax]. Oxford: OUP.
- Kiss, Katalina É. 2002. *The Syntax of Hungarian* [Cambridge Syntax Guides]. Cambridge: CUP. doi:10.1017/CBO9780511755088
- Kiss, Katalina É. 2006. The function and syntax of the verbal particle. In *Event Structure and the Left Periphery – Studies on Hungarian*, Katalina É. Kiss (ed.), 17–56. Dordrecht: Springer. doi:10.1007/978-1-4020-4755-8_2
- Klein, Wolfgang. 1994. *Time and Language*. London: Routledge.
- Krifka, Manfred. 1998. The origins of telicity. In *Events and Grammar*, Susan Rothstein (ed.), 197–236. Dordrecht: Kluwer. doi:10.1007/978-94-011-3969-4_9
- Leinonen, Marja. 1982. *Russian Aspect, 'temporal'naja lokalizacija', and Definiteness/ Indefiniteness*. PhD dissertation, Helsinki University.
- Maslov, Jurij. 1958. Rol' tak nazyvaemoi perfektiivacii i imperfektivacii v processe vozniknovenia glagol'nogo vida (The Role of So-called Perfectivisation and Imperfectivisation in the Process of Appearance of Verbal Aspect). *Meždunarodny s'ezd slavistov. Doklady*: 3–39. Moskva: Izdatel'stvo Akademii nauk SSSR.
- McCawley, James. 1971. Tense and time reference in English. In *Studies in Linguistic Semantics*, Charles J. Fillmore & D. Terence Langendoen (eds), 97–114. New York NY: Holt, Rinehart and Winston.
- Michaelis, Laura A. 2004. Type shifting in Construction Grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15: 1–67. doi:10.1515/cogl.2004.001
- Michaelis, Laura A. 2016. Serving a slice of time: Unifying tense, aspect and modality with aspectual Constructions. Paper presented at the Conference Tense, Aspect, Modality, Evidentiality: Comparative, Cognitive, Theoretical, Applied Perspectives, Université Paris Diderot.
- Moens, Marc & Steedman, Mark. 1988. Temporal ontology and temporal reference. In *The Language of Time: A Reader* [2005], Inderjeet Mani, James Pustejovsky & Robert Gaizauskas (eds), 93–114. Oxford: OUP.
- Padučeva, Elena V. 1996. *Semantičeskie Issedovanija*. Moscow: Škola 'Jazyki ruskoj kul'tury.
- Piñon, Christopher. 2006. Weak and strong accomplishments. In *Event Structure and the left periphery – Studies on Hungarian*, Katalina É. Kiss (ed.), 91–106. Dordrecht: Springer. doi:10.1007/978-1-4020-4755-8_5
- Plungian, Vladimir. 2012. Predislovie. Tipologija aspektual'nyx system i kategorij (Introduction. Typology of aspectual systems and categories). In *Issledovanija po teorii grammatiki* (Research into Theories of Grammar), Nikolaj N. Kazanskij (ed.), 7–42. Sankt-Peterburg: Nauka.

- Ramchand, Gillian. 2004. Time and the event: The semantics of Russian prefixes. *Nordlyd* 32(2): 323–361.
- Sartre, Jean-Paul. 1947. Explication de l'étranger. *Situations I*. Paris: Gallimard.
- van Schooneveld, Cornelis Hendrik. 1978. *Semantic Transmutations: Prolegomena to a Calculus of Meaning*. Bloomington IN: Phylaxardt.
- Smith, Carlota. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer. doi:10.1007/978-94-015-7911-7
- Squartini, Mario & Bertinetto, Pier Marco. 2000. The simple and compound past in Romance languages. In *Tense and Aspect in the Languages of Europe*, Östen Dahl (ed.), 403–439. Berlin: Mouton de Gruyter.
- Svenonius, Peter. 2004. Slavic prefixes inside and outside VP. *Nordlyd* 32(2): 205–253.
- de Swart, Henriëtte. 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory* 16: 347–385. doi:10.1023/A:1005916004600
- de Swart, Henriëtte. 2000. Tense, aspect and coercion in a cross-linguistic perspective. In *Proceedings of the Berkeley Formal Grammar Conference*, Miriam Butt & Tracy Holloway King (eds). Stanford CA: CSLI.
- de Swart, Henriëtte & Molendijk, Arie. 2002. Le passé composé narratif: Une analyse discursive de L'Étranger de Camus. In *Temps et aspect – De la morphologie à l'interprétation*, Brenda Laca (ed.), 193–215. Saint Denis: Presses Universitaires de Vincennes.
- Talmy, Leonard. 2000. *Towards a Cognitive Semantics*, Vol. 2: *Typology and Process in Concept Structuring*. Cambridge MA: The MIT Press.
- Tenny, Carol. 1994. *Aspectual Roles and the Syntax-Semantics Interface*. Dordrecht: Kluwer. doi:10.1007/978-94-011-1150-8
- Tournadre, Nicolas. 2004. Typologie des aspects verbaux et intégration à une théorie du TAM. *Bulletin de la Société Linguistique de Paris* XCIV(1): 7–68. doi:10.2143/BSL.99.1.541909
- Verkuyl, Henk J. 1989. Aspectual classes and aspectual composition. *Linguistics and Philosophy* 12: 39–94. doi:10.1007/BF00627398
- Vet, Co. 1980. *Temps, aspects et adverbes de temps en français contemporain*. Genève: Droz.
- Zaliznjak, Anna & Šmelev, Aleksej. 1997. *Lekcii po ruskoi aspektologii* (Lectures on Russian Aspectology). Munich: Otto Sagner.

Phraseological usage patterns of past tenses

A corpus-driven look on French *passé composé* and *imparfait*

Oliver Wicher

Paderborn University

This chapter presents a corpus-driven investigation into usage patterns of the French past tenses *passé composé* and *imparfait*. Using a new reference corpus of French, the *Corpus de référence du français contemporain*, and adopting a construction grammar perspective, we analyze past-tense occurrences of two highly frequent polysemous verbs, *vouloir* ‘want’ and *voir* ‘see’. Assuming that the two tenses are an alternation phenomenon, collocations can be identified that differ in terms of their attracted complements. These highly distinctive collexemes indicate phraseological uses of past-tensed verbs whose tense choice is constrained. The results provide further evidence for the inseparability of lexis and grammar and the semantic-pragmatic shifts that are prevalent in past-tense constructions of highly frequent polysemous verbs.

Keywords: *passé composé*, *imparfait*, corpus-driven, lexical grammar, collocational analysis

1. Introduction

Construction Grammar (CG) theories have been a popular approach popular approach to language description and acquisition, as they offer a cognitive account of linguistic structures, overcoming the strict division between lexis and grammar postulated in generative frameworks. One of the most important findings has been that speakers resort to a vast array of more or less schematized multi-word units that can be analyzed as conventionalized form-meaning correspondences, so-called constructions. It must be noted, however, that the body of CG research has mainly focused on the analysis of English. Approaches to French are still in a nascent stage, although the Romance languages have recently received increasing attention (cf. the volumes edited by Bouveret & Legallois 2012; De Knop et al. 2013; Boas & González-García 2014; Legallois & Patard 2017). A similar situation

can be found for corpus linguistics, where most work draws on written-language corpora such as FRANTEXT (ATILF – CNRS & Université de Lorraine 2016) or smaller spoken-language corpora such as ESLO (Eshkol-Taravella et al. 2011). A genre-diverse reference corpus and a corpus-based descriptive grammar have been urgent desiderata for a long time (Deulofeu & Debaisieux 2012; Bilger & Cappeau 2013; for a critical overview of existing French corpora cf. Siepmann et al. 2017). In addition, the majority of corpus research on French seems to be qualitative in nature. Hypotheses are checked against attested data, but rarely by using inferential statistics (cf. the points of criticism expressed in contributions such as Gries 2010).

The French tenses *passé composé* (PC) and *imparfait* (IMP) have been thoroughly investigated, but mostly from the point of view of theoretical linguistics, especially in structuralist-functionalist frameworks, and of first and second language acquisition. Existing corpus studies are often limited to the description of overall tense frequencies (see Section 2.1). One question that has not yet been tackled from a quantitative corpus perspective is the interaction of past-tensed verbs and their lexico-grammatical environments. The rationale behind this is the consideration that past-tense choice in French correlates with the lexico-grammatical environment of the verb: if the tenses represent grammatical aspect, then factors like the semantics of the complement or temporal adverbs come into play. The tenses can also express temporal and modal values and certain interpretations are lexico-syntactically constrained (e.g. counterfactuality in binary structures such as *Une minute de plus, le train déraillait* ‘One more minute and the train would have derailed’; cf. Berthonneau & Kleiber 2006: 8). It thus seems plausible to explore past-tense choice from a lexico-grammatical corpus perspective.

Corpus-based constructionist work has generated substantial evidence for the inseparability of lexis and grammar (Römer 2009: 140), giving rise to a diversity of statistical methods that measure the association strength of linguistic units to each other. One of these methods is collostructional analysis (Stefanowitsch & Gries 2003 and subsequent work), which identifies lexemes that are distinctive for a construction (hence the notion of ‘collostruction’, a blend of collocation and construction).

I would like to argue that collostructional analysis permits to extract highly distinctive complements of past-tense constructions such as the most strongly attracted verbs of ‘*je voulais / j’ai voulu + V*’ or distinctive noun phrase complements of ‘*je voyais / j’ai vu + NP*’. These distinctive collexemes are in turn indicative of phraseological patterns. It is to be expected that tense choice in highly entrenched patterns differs from the traditional tense behaviour of a verb, since phraseological patterns may conventionalize, obtain new semantic and pragmatic values, and overshadow the initial meaning (Bybee 2006). The following research questions will thus be addressed:

- a. If collostructional analysis is applied to the past-tense constructions of a French verb, to what extent do the PC- and the IMP-constructions differ from one another in their distinctive collexemes?
- b. How can tense choice in phraseological uses of past-tensed verbs be explained with the co-occurrence of distinctive collexemes?

These questions will be addressed by introducing a new reference corpus of contemporary French, the *Corpus de référence du français contemporain* (CRFC) (Siepmann et al. 2017), and analyzing two examples of highly frequent polysemous verbs: the modal *vouloir* ‘want’ and the perception-/cognition-verb *voir* ‘see’. This chapter is organized as follows: Section 2 sets the theoretical framework, discussing previous work on PC and IMP and exploring how corpora and constructions can offer a fresh view on their analysis. Section 3 introduces the CRFC and the methodology of the corpus study. Section 4 presents the two case studies on French past-tense constructions and explores their usage patterns. Section 5 discusses the findings and the theoretical implications for a constructional analysis of past-tensed verbs. Section 6 draws a brief conclusion.

2. Theoretical framework

2.1 French past tenses between theoretical and corpus linguistics

The plethora of studies on French past tenses may be divided into three general areas: (a) theoretical linguistics; (b) corpus linguistics; and (c) language acquisition (e.g. Ayoun 2013). Theoretical studies are mainly concerned with the semantic values of the past tenses. First and foremost, this concerns the notion of aspect, that is, the internal temporal constituency of a situation (Comrie 1976: 3).¹ It is commonly assumed that the distinction between PC and IMP reflects the opposition between perfective and imperfective aspects: the former presents a situation as bounded, whereas the latter construes it as unbounded (Sten 1952; Garey 1957; Imbs 1960; Klum 1961; Vettters 1996). What distinguishes the PC from the equally bounded *passé simple* is its connection to speech time and thus the possibility to encode resultative aspect (Wilmet 1997; Luscher 1998; de Saussure 2000; Vettters &

1. Due to space constraints, we cannot delve into the terminological confusion of aspectology and into the controversy over mono- vs. bidimensional approaches to aspectuality (but cf. Sasse 2002 for an exhaustive overview of these issues). We operate with the notions *lexical aspect*, describing the inherent event semantics of the predicate, and *grammatical aspect*, encoded by past-tense choice (cf. also the respective contributions in the recent state-of-the-art volume edited by Binnick 2012).

de Mulder 2000 among others; for discussion see Molendijk 2002). The aspectual distinction is also relevant for the structuring of discourse, as the perfective tenses foreground events in narrative, whereas the IMP creates periods of backgrounding and stasis (Weinrich 1964; Hopper 1979; Kamp & Rohrer 1983; Fleischman 1985).

Generally speaking, past-tense choice belongs to the domain of grammatical aspect and has often been associated with the notion of viewpoint (Smith 1991), as it depends on the speaker's perspective on an event. Grammatical aspect is opposed to lexical aspect, the event structure implied by the lexical semantics of the verb. As for the interaction between these two aspectual categories, every verb can be realized in both tenses, but stative verbs show a preference for the IMP while inherently telic verbs tend to prefer the PC (Boogaart 2004: 1178–1179). However, coercion effects can be triggered if semantic conflicts arise (De Swart 1998). Telic verbs such as *atteindre* 'reach' in (1a) denote the pre-ultimate phase before event completion if used with the IMP in (1b). A habitual reading may be produced as well if the context allows for such an interpretation. Coercion effects can also be an indicator of meaning shifts: a stative verb such as *savoir* 'know' in (2a) is telicized if used with the PC, receiving a bounded construal in (2b).

- (1) a. *J'ai atteint le sommet de la montagne.*
'I reached / I have reached the mountain top'.
b. *J'atteignais le sommet de la montagne.*
'I was reaching / I used to reach the mountain top'.
- (2) a. *Je pensais qu'il se tapait Emilie! Et donc il savait toute la vérité!* (TV)
'I thought he was banging Emilie! So he knew the full truth!'
b. *Parce que moi aussi je t'ai trompé, dès que j'ai su la vérité sur toi.* (Fiction)
'Because I also cheated on you, as soon as I found out the truth about you'.

Boundedness may not only be modified via past-tense choice, but can also be manipulated by other sentential elements. A quantized object noun triggers a telic interpretation of the event like in (3a), whereas an indefinite quantity detelicizes the predicate in (3b) (Carlson 1984; Krifka 1989). The verb *nager* 'swim' in (4a) is considered an atelic activity, but combined with a directional prepositional phrase as in (4b), it receives a telic reading. In view of these few examples, it should be clear that many verbs cannot invariably be related to one aspectual class, but rather receive a specific construal depending on their lexico-syntactic environment.

- (3) a. *Je mange trois pommes.*
'I eat three apples'.
b. *Je mange des pommes.*
'I eat apples'.

- (4) a. *Je nage.*
 'I swim'.
 b. *Je nage jusqu'au bord de la rivière.*
 'I swim up to the river bank'.

The distinction between these two tenses is not only a matter of aspectual, but also of temporal values. Both of them may be prototypically called past tenses, but they can equally refer to the present or to the future (cf. Vet 1992; Larrivée 2002; Desclés & Guentchéva 2003 or Antoniou 2005 for discussions of contextual interpretations of the PC). Consider, for instance, (1a), where the event of having reached the mountain top may be entirely located in the past, or coincide with the moment of utterance. The same goes for the IMP as it encompasses a variety of semantic values, many of which have been subject to lively controversy such as the narrative (Labeau 2005), the counterfactual (Berthonneau & Kleiber 2006), the pre-ludic (Patard 2010) or the attenuative use (Berthonneau & Kleiber 1994; cf. Labeau 2002 and Patard 2007 for an overview of IMP semantics). In short, both tenses are highly polysemous, representing various temporal (past, present and future temporality), aspectual or modal meanings.

Against the background of these considerations, (quantitative) corpus approaches seem particularly promising. Not only do corpora provide data to test hypotheses against actual usage instances, a corpus-driven approach also permits us to make deductions about the semantics of a structure from its frequencies and distributions of (co-)occurrences (Tognini-Bonelli 2001; Stubbs 2015). This approach may help to disambiguate the polysemy of the PC and IMP, as certain functions are reflected in typical lexico-grammatical patterns (see e.g. Detges 2010 on the correlation between the attenuative IMP and the '*je voulais* + INF' pattern). Yet the number of corpus studies is somewhat modest. The most commonly investigated genres are oral narrations (Labelle 1987), newspaper articles (Herzog 1981; Waugh & Monville-Burston 1986; Facques 2002), television talk (Labeau 2006), sports commentaries (Labeau 2004, 2007) and obituaries (Do-Hurinville 2010; Labeau 2009, 2013). Several problems emerge, however, if one wanted to make a generalization about the findings of past-tense distributions. First, the corpora are rather restricted in terms of their size and therefore their representativeness, covering few volumes of journal articles or manual transcriptions of TV shows. To the best of our knowledge, the only studies making explicit use of genre-diverse (self-compiled) corpora are Di Vito (1997) with 53,000 clauses distributed in spoken (television talks, interviews) and written (fiction, detective novels, magazines) genres; Rebotier's (2005) contrastive analysis with about 1,640 French and German verbs in fiction, theatre scripts and popular scientific literature; and Patard's (2007) investigation into the IMP with about 4,400 occurrences in literary, journalistic and conversational data.

Second, the studies mostly provide raw or percentage frequencies; thus, apart from Rebotier's study, inferential statistics that could help to generalize the findings are not used. Third, the studies are restricted to analyzing verbs and fail to take into account the characteristic usage patterns in which they may occur. Accordingly, they do not quantify whether verbs show preferences for one particular past tense nor do they analyze whether there are recurrent strings that hint at lexically-specific past-tense constructions with tense constraints.

Given these caveats, a quantitative corpus-driven approach may be difficult to implement if the corpus exhibits a genre bias and is limited in size. However, this lacuna can be remedied for contemporary French with the CRFC (see Section 3). But let us first see how a corpus-driven constructional approach can offer a fresh view on French verbs and their past-tense occurrences.

2.2 Construction grammar, complementation, and the analysis of French past tenses

CG approaches to language description have proven to be an interesting alternative to mainstream generative theories, since they account for structures such as multi-word expressions whose frequently idiosyncratic syntax cannot be motivated by transformations as in generative grammar. Instead, the central linguistic unit posited is a conventionalized form-meaning correspondence, i.e. a construction. As such, it resembles the Saussurean sign, but extends to form-meaning pairings at the higher levels of abstraction like argument structure constructions or whole clauses. Whereas in earlier work (e.g. Goldberg 1995) the non-compositionality of a construction was its central defining criterion,² the scope has now been extended to usage: as long as a linguistic unit occurs with sufficient frequency, it can also be considered a construction, even if its meaning is fully compositional (Goldberg 2006).

One major assumption is the inseparability of lexis and grammar (Römer 2009): constructions can be more or less schematized and are located on a syntax-lexicon continuum (Croft & Cruse 2004: 255; Broccias 2012). At the one end of the continuum, they form an abstract schema such as the ditransitive construction which expresses an unpredictable sense of transfer (Goldberg 1995), but which also occurs preferably with a discrete set of verbs. At the other end of the continuum, there are frozen expressions such as idiomatic expressions. It is therefore not surprising that CG is closely connected to phraseology, which in its widest sense is concerned with recurrent co-occurrences of linguistic elements, be they lexical or grammatical (Gries 2008: 5). Finally, CG provides a cognitive account of language and language

2. The reader may recall pioneering studies on constructions such as 'let alone' (Fillmore et al. 1988) or 'What's X doing Y?' (Kay & Fillmore 1999).

acquisition as it is highly compatible with usage-based approaches (Diessel 2015; Ellis et al. 2016). Frequency effects influence from the very beginning the acquisition of constructions; highly frequent exemplars are learned and stored first, priming the acquisition of less frequent members. They are also prone to be entrenched as autonomous constructions and may develop new semantic-pragmatic interpretations that override the source meaning (see e.g. Bybee 2006 and her analysis of the English *going to*-construction that grammaticalized from a purpose construction to a future marker).

What can corpus-driven CG add to the existing body of research on French verbs and their past tenses? Without being able to postulate the exact number and network of past-tense constructions, which would go beyond the scope of this chapter, it seems *a priori* plausible to consider PC and IMP as genuine constructions. As sketched in Section 2.1, these tenses are characterized by form-meaning asymmetries because their concrete semantic value is often not directly predictable from the construction itself. For instance, Detges (2010) argues that the IMP instantiates an abstract background construction out of which has emerged the polite imperfect. Similarly, Michaelis (1998) identifies several constructions and their network relationships in order to explain the usages of the English present perfect. What this chapter focusses on, however, is the interdependencies between lexis and grammar of French past-tensed verbs. In order to explore their lexical grammar, collostructional analysis is applied.

The basic procedure is as follows: a given grammatical construction has an open verb slot such as the ditransitive [SUBJ V OBJ₁ OBJ₂]. Many verbs can occur in the slot, but collostructional analysis reveals which verbs are in fact most strongly attracted to it. This is done by considering observed frequencies of (co-)occurrences and by calculating expected frequencies. Submitting the data to a non-parametric contingency test such as the Fisher-Yates-Test, one can calculate the association strength between the construction and its collexemes, the so-called collostructional strength. The collexemes are ranked and may be clustered. The method has been primarily developed for the analysis of individual constructions and their most frequent collexemes, but it is also possible to perform a distinctive collexeme analysis and to investigate to what extent two near-synonymous constructions differ with regard to their distinctive collexemes (e.g. ditransitive vs. prepositional dative; Gries & Stefanowitsch 2004).

The idea is now to treat the PC- and the IMP-constructions as if they were near-synonymous and therefore an alternation. The rationale behind this is the consideration that both constructions prototypically locate an event in the past, but with different perspectives; further alternations are added because both tenses can also encode present or future temporality. If there are distinctive collexemes of one of the French past-tense constructions, that is, collocates that appear with

significant frequency, then it is reasonable to assume that there is a correlation between the semantics of this collexeme and the past-tense construction. Tense choice is expected to be constrained in these cases. Note that in this regard, collostructional analysis is not performed in its traditional top-down way, which would mean to consider the verbal heads as collexemes, but with respect to individual past-tense constructions of verbs.

The collexemes in question are the complements of the respective past-tense constructions, which can be motivated by their pivotal role in the syntax and semantics of verbs. In this regard, a suitable framework is the constructional verb typology of French developed by Willems (2012). Based on work of predecessors such as Gross (1985), its point of departure is the observation that verbs sharing the same syntax also have common semantics. Consider, for example, meteorological verbs, all of which are avalent and realized with an impersonal subject (*il neige* ‘it snows’, *il pleut* ‘it rains’, etc.) or communication verbs that share the ditransitive pattern (*demander* ‘ask’, *dire* ‘say’, etc.).

Briefly put, three macro-constructions (nominal, infinitival, clausal) are postulated in a first step, depending on the complementation type. These constructions are then each divided into semantic-aspectual classes (states, activities, processes and achievements, all of which can also have causative variants with valency augmentation).

Finally, valency is mapped upon these aspectual classes, resulting in a fine-grained classification to which over 7000 French verbs are assigned. Since many verbs display several different syntactic configurations, they can be assigned to more than one group.

The case studies presented in this chapter deal with two examples of well-known verbs. The first case is the modal *vouloir* ‘want’, which, as a stative verb, tends to be encoded with the IMP. However, two main usages must be distinguished. First, the construction that expresses volition where past-tense usage is said to be a matter of temporal distance: the PC refers to a direct temporal connection between the subject’s wish and its realization as in (5a), whereas the IMP suggests some vague future as in (5b) (Blumenthal 1976: 36). Second, the speech act-construction that exclusively licenses the IMP and combines with an infinitive expressing an attenuation as in (5c) (Berthonneau & Kleiber 1994; Abouda 2004; Anscombe 2004; Detges 2010).

- (5) a. *Ce mineur... a voulu que ses quatre enfants fassent des études. Ils en ont fait.*
 ‘This miner... wanted his four children to study. They did so’.
- b. *Je voulais être instituteur.* (Blumenthal 1976: 36)
 ‘I wanted to become a (primary) teacher’.
- c. *Je voulais te demander si tu étais dispo.*
 ‘I wanted to ask you if you were available’.

The second case *voir* ‘see’ belongs to a class of polysemous verbs with semi-regular meaning extensions (Willems 2012: 42), allowing a perceptual interpretation (*je vois une ombre* ‘I see a shadow’ or *je vois arriver le prof* ‘I see the teacher arriving’) and a cognitive one (*je vois ce que tu veux dire* ‘I see what you mean’). It is also fairly flexible in its complementation and has certain grammaticalized usage patterns as a semi-auxiliary, for instance in a pseudo-passive construction such as *je me vois partir* ‘I see myself leaving’ (cf. Chocheyras 1968, Bat-Zeev Shyldkrot 1984, Labelle 1996). Table 1 gives an overview of the different complementation patterns according to Willems’ (2012) typology.

Table 1. Complementation patterns of *vouloir* and *see*

| | vouloir ‘want’ | voir ‘see’ |
|-------------|-------------------------------------------|---------------------------------------|
| nominal | vouloir + NP | voir + NP |
| infinitival | vouloir + INF | voir + INF + NP |
| clausal | vouloir + que + sub. clause ‘want that..’ | voir + que + sub. clause ‘see that..’ |

If Willems’ typology captures constructional polysemy, then it may be reasonable to presume that past-tense usage may also be correlated to one of the complementation patterns of these verbs (cf. Hare et al. 2003 for psycholinguistic evidence of the link between verb sense and complementation). Although it would probably be premature to postulate a causal relationship between a complementation pattern and past-tense choice, the typology can serve as a framework in order to systematically compare usage patterns of French past-tensed verbs. The following Section 3 will present the reference corpus and the methodology of the empirical study.

3. Methodology: Corpus and data analysis

The CRFC is a genre-diverse reference corpus of contemporary French (Siepmann et al. 2017).³ It includes about 310 million words, evenly distributed among spoken, written and pseudo-spoken varieties. Table 2 illustrates its composition.

3. Initiated under direction of Dirk Siepmann (University of Osnabrück) and Christoph Bürgel (Paderborn University), in collaboration with Sascha Diwersy (University of Montpellier). The author has also been also a member of the research group since 2017.

Table 2. Composition of the CRFC (adapted from Siepmann et al. 2017: 70)

| Category | Subcorpus | Size in million | |
|---------------|----------------------|-----------------|-----------|
| Spoken | Informal | 30 | |
| | Formal | 30 | |
| Pseudo-spoken | Drama scripts | 30 | total 155 |
| | Television subtitles | 2.5 | |
| | Messages/chats | 2.5 | |
| | Discussion forums | 60 | |
| Written | Academic | 30 | |
| | Non-academic | 30 | |
| | Prose fiction | 30 | |
| | Newspapers | 45 | total 155 |
| | Magazines | 10 | |
| | Diaries/blogs | 5 | |
| | Letters/e-mails | 1 | |
| | Miscellaneous | 4 | |

Table 2 shows that 25% of the spoken informal subcorpus consists of several CC-licensed corpora. The remaining other 75% is made up of unscripted television talk of a large variety of different television shows, totaling 30 million words. Note that these data have been obtained in the form of subtitles. Several studies have shown that they are a viable solution for analyzing actual speech, at least in the domain of lexical grammar (e.g. New et al. 2007; Quaglio 2009; Levshina 2017). The spoken formal subcorpus consists of parliamentary speeches and academic lectures, totaling another 30 million words. The written subcorpora cover a large range of different genres such as prose fiction, academic texts and newspapers, totaling 155 million words. Finally, the pseudo-spoken subcorpus – written language that exhibits typically spoken language characteristics – includes television subtitles, discussion forums, drama scripts, chats (Falaise 2005) and text messages (Panckhurst et al. 2013), totaling 95 million words.

The CRFC currently lacks syntactic parsing, but has been POS-tagged using French Tree Tagger (Stein 2003). A monitor corpus will be set up in order to capture linguistic innovations. The corpus is currently available on-demand on Sketch Engine. Several studies have already been conducted using the CRFC. Siepmann (2015) demonstrates that it captures frequencies and patterns of colloquial lexemes such as *foutre*, which are barely accounted for by other spoken corpora. Siepmann and Bürgel (2015) show that the use of the subjunctif is restricted to a small number of lexical and phrasal triggers. Furthermore, they investigate the use of French prepositions from a corpus-driven perspective (Siepmann & Bürgel 2016). Inspired by the pioneering work of Martinez & Schmitt (2012), the authors have developed

frequency-based phrasal expressions lists for spoken and written French (Siepmann & Bürgel to appear).

Collostructional analysis was performed in order to determine which complements are distinctive of the respective past-tense constructions. Two levels of analysis were considered. First, a distinctive collexeme analysis was carried out. The statistical calculation led to lists of collexemes that are distinctive for the perfective and the imperfective construction of the two verbs, respectively. Second, concordances of the occurrences in which these collexemes occur were analyzed from a more qualitative perspective: what are highly frequent phraseological patterns and how to they correlate with tense choice?

To extract the relevant past-tense constructions, the most schematic construction was labeled as a tripartite structure [SUBJ V_{PST} COMP], where the verbal slot was either the periphrastic PC or the synthetic IMP. The complementation slot could be, based on Willems' typology, a noun phrase, a verbal phrase, or the complementizer *que* 'that'. Using the corresponding CQL-syntax commands, it was possible to retrieve the relevant instances. To improve recall, additional slots were taken into account for optional elements like adverbs (*j'ai bien vu* 'I clearly saw') or clitics (*je voulais vous en parler* 'I wanted to talk to you about this').⁴ As for the IMP-constructions, the conjunction *si/s'* 'if' in the left periphery was excluded in order to avoid irrealis conditional sentences, where the IMP would be automatically triggered by the conditional construction (e.g. *si tu voyais* 'if you saw'). The most frequent complements were exhaustively listed for the PC- and the IMP-construction until a minimum frequency of 5.

The collostructional analysis itself was performed using the R script *Coll. Analysis* 3.2a (Gries 2007). The distinctive collexemes in terms were listed of their collostructional strength, i.e. the log-transformed *p*-values of the one-tailed exact Fisher-Yates-Test. Due to space restrictions, only the 15 most distinctive ones are presented if there were more. The significance value was set at a collostructional strength of greater than 3, corresponding to *p* < 0.001. Such a high significance level can be motivated by the nature of the past-tense construction: its complementation slot is entirely open; in a construction such as 'vouloir_{PST} + V' any verb can theoretically occur in the slot. This in turn means that only few collexemes exhibit an extraordinary collostructional strength.

4. The CQL-commands were as follows:

[lemma = "avoir" & tag = "VER:pres"] []? [lemma = "vouloir|voir" & tag = "VER:pper"] for the perfective construction, permitting to retrieve plural and/or feminine past participals as well; [lemma = "vouloir|voir" & tag = "VER:impf"] for the imperfective construction, with negative filter on *si/s'* from -5L to 0. The nominal complements were obtained by searching for the string of tags DET:ART or PRO:DEM followed by NOM, e.g. *la guerre* 'the war' or *cette guerre* 'that war'.

4. Results

4.1 Past-tense choice at the semantics-pragmatics interface: The modal verb *vouloir*

This section presents the results of the collostructional analysis and the concordancing of distinctive collexemes. For the case study of *vouloir*, the CRFC provides 18,300 PC- and 44,974 IMP-constructions. Both of them mostly attract verbal complements and differ sharply in terms of distinctive collexemes (see Table 3). Although highly significant, no distinctive PC-collexeme has a collostructional strength greater than 16. By contrast, some collexemes are exceptionally strongly attracted to the IMP.

Table 3. Distinctive collexemes of the past-tense constructions of *vouloir* (Coll.Str. > 3)

| PC with nominal complementation (N = 546) | | IMP with nominal complementation (N = 1,759) | |
|----------------------------------------------------|----------|-----------------------------------------------------|----------|
| Collexeme | Coll.str | Collexeme | Coll.str |
| <i>ouverture</i> ('opening') | 5.94 | <i>enfant</i> ('child') | 3.93 |
| <i>guerre</i> ('war') | 5.30 | | |
| <i>législateur</i> ('legislator') | 5.02 | | |
| <i>réforme</i> ('reform') | 4.71 | | |
| PC with verbal complementation (N = 16,489) | | PC with verbal complementation (N = 39,140) | |
| <i>prendre</i> ('take') | 15.92 | <i>dire</i> ('say') | 265.00 |
| <i>prêter</i> ('lend') | 15.60 | <i>savoir</i> ('know') | 154.37 |
| <i>donner</i> ('give') | 15.14 | <i>demander</i> ('ask') | 54.87 |
| <i>mettre</i> ('put') | 14.60 | <i>parler</i> ('talk') | 48.83 |
| <i>reprendre</i> ('take back') | 11.10 | <i>être</i> ('be') | 38.86 |
| <i>créer</i> ('create') | 10.85 | <i>remercier</i> ('thank') | 19.76 |
| <i>répondre</i> ('answer') | 9.82 | <i>avoir</i> ('have') | 14.10 |
| <i>croire</i> ('believe') | 9.08 | <i>voir</i> ('see') | 12.06 |
| <i>montrer</i> ('show') | 7.60 | <i>rester</i> ('stay') | 8.63 |
| <i>comprendre</i> ('understand') | 5.66 | <i>vivre</i> ('live') | 8.48 |
| <i>faire</i> ('make, do') | 4.91 | <i>devenir</i> ('become') | 7.71 |
| <i>essayer</i> ('try') | 4.71 | <i>entendre</i> ('hear') | 6.67 |
| <i>rendre</i> ('return, give back') | 4.65 | <i>venir</i> ('come') | 3.47 |
| <i>jouer</i> ('play') | 3.64 | | |
| <i>tuer</i> ('kill') | 3.43 | | |
| 1 other | | | |
| PC with <i>que</i> -complementation (N = 1,265) | | IMP with <i>que</i> -complementation (N = 4,075) | |
| | | <i>que</i> ('that') | 9.45 |

There are few nominal collexemes and they seem to be related to the field of formal-administrative language for the PC. The only IMP-collexeme is *enfant* ‘child’. They may be significantly attracted to the respective construction, but the past-tense occurrences can hardly be considered phraseological in these cases as shown in (6)–(8).

- (6) *le Premier ministre d'un autre pays est le chef de sa majorité, le chef de l'État doit rassembler tous les Français.* </p> <p> *C'est pour cela que j'ai voulu l'ouverture.*

(Spoken Formal)

‘The prime minister of another country is the leader of its majority, the Head of State must bring together all French people. </p> <p> That is why I wanted the opening.’

- (7) *depuis 1948 combien de morts israéliens et combien de morts palestiniens? ils sont chez eux...les palestiniens n'ont jamais voulu cette guerre...*

(Discussion Forums)

‘since 1948 how many Israeli deaths and how many Palestinian deaths? They are at home... the Palestinians have never wanted this war..’

- (8) *PÈRE: (...). Ma femme voulait des enfants, une grande famille. Je lui en ai fait huit.* (Drama)

‘FATHER: (...) My wife wanted children, a big family. I gave her eight [children].’

Instead, most interesting are the verbal complements at which a closer look will now be taken. Distinctive PC-collexemes are verbs like *prendre* ‘take’, *prêter* ‘lend’, *donner* ‘give’ and *mettre* ‘put’. It is difficult to identify a common semantic trait to these verbs, apart from their transitivity. By contrast, the IMP-construction has several highly distinctive collexemes, namely *dire* ‘say’, *savoir* ‘know’ *demander* ‘ask’ and *parler* ‘talk’. A closer look at these and the other IMP-collexemes reveals a striking pattern: at least five verbs are clearly related to the use of the polite imperfect. A randomized concordance displays patterns such as *je voulais dire que* ‘I wanted to say that’, *je voulais savoir si* ‘I wanted to know if’, *je voulais te/vous demander* ‘I wanted to ask you’ or *je voulais vous remercier* ‘I wanted to thank you’ (see Figure 1). Examining the preferred pronouns that occur in these constructions, one might even go as far as to postulate inflectional islands (Rice & Newman 2005), since they almost exclusively choose 1SG as the subject and 2SG/PL as the object.

The crucial point is that none of these collexemes occurs significantly with the alternative PC-construction. This distribution provides quantitative evidence for the theoretical claim that, as sketched in Section 2.2, the polite imperfect cannot occur with the PC. Counterexamples such as *j'ai voulu savoir* or *j'ai voulu dire* are very rare and express literal volition as in (9) and (10). Moreover, a certain correlation may be cautiously established between the use of perfective *vouloir*, narrative frames, temporal proximity and transitivity. Consider for example (11): the speaker

| | | | |
|--------|--------------------------------------|----------------|--------------------------------------|
| Forums | n'est pas d'actualité ?? Bonjour, je | <i>voulais</i> | juste dire que j' étais entièrement |
| Drama | pas ? ... Joyce Lee à l'appareil, je | <i>voulais</i> | juste vous dire que... Sami lui |
| Drama | vous méprenez pas Pénélope ... Je | <i>voulais</i> | dire ... Soyez ma maitresse, dure |
| Film | aurais pas laissés partir. Julie. Je | <i>voulais</i> | te dire... enfin ... Juste un petit |
| TV | donc trôner. – A. Ducret : Moi, je | <i>voulais</i> | vous dire, le sketch, on vous le |
| Forums | nom de l'hôpital c'est gentil cc je | <i>voulais</i> | te dire nancy ms j'ai vu plus haut |
| Drama | du village qui s'éteint ... Mais tu | <i>voulais</i> | me parler, je suppose. Juliette Oui |
| Drama | Promis. A 22H on dort ... Et toi tu | <i>voulais</i> | me parler ? La Guéramine |
| Film | Tu veux me rendre jaloux ? Elle | <i>voulait</i> | te parler. Me vider de tout mon. |
| TV | facilement Voici le kit dont je | <i>voulais</i> | vous parler. C'est un problème de |
| TV | avec une grosse actualité. Je | <i>voulais</i> | vous remercier. Quelle chance |
| TV | ma réponse sur ... – Le prix. – Je | <i>voulais</i> | savoir si... – La négo, ce n'est pas |
| Forums | de ça je suis pas fatiguée.. dc je | <i>voulais</i> | savoir si ça pouvait qd meme |
| Forums | ma soupe 804 demain et je | <i>voulais</i> | savoir quelle base de légumes et |

Figure 1. Concordance of the IMP-construction of *vouloir* + V

tells the story of how he wanted to save his friend's guinea pigs which were to be killed. This act of wanting to save them is foregrounded and must be realized perfectly in order to make the narrative advance. The grammatical correlate to this pattern is perfective *vouloir* combined with a dynamic verbal phrase.

- (9) *C'est ainsi que j'ai croisé le regard brillant de Luis, originaire du Honduras. Et j'ai voulu savoir quels étaient les réels risques de ce voyage.* (Spoken Informal)
 'That's how I crossed the brilliant eyes of Luis, who originally came from Honduras. And I wanted to know what were the real risks of this trip'.
- (10) *Georges: – Non! Ce n'est pas ce que j'ai voulu dire.* (Drama)
 'Georges: No! This isn't what I wanted to say'.
- (11) *Le bémol de mon Noël: j'ai appris qu'une amie a donné 15 bebes hamsters a quelqu'un pour qu'il les tue. (...) Elle a dû se dire qu'ils n'étaient QUE des hamsters. J'ai voulu les sauver mais il était déjà trop tard.* (Discussion forums)
 'The B-flat of my Christmas: I learned that a friend gave 15 baby guinea pigs to somebody to have them killed. She must have said to herself that they were ONLY guinea pigs. I wanted to save them, but it was already too late'.

It thus appears that the lion's share of past-tensed *vouloir*-constructions is instantiated by verbal complements, whereas nouns and clausal *que*-complementation only make up for a small part of the overall occurrences. Perfective *vouloir* combines with dynamic transitive verbs representing foregrounded events in discourse, whereas imperfective *vouloir* predominantly instantiates the polite imperfect. These results give further empirical evidence for Detges' (2010) claim that it is in first place entrenched in the speakers' mental corpus (Taylor 2012) and not computed.

4.2 Between cognition and perception: The verb *voir*

Regardless of the complementation type, *voir* occurs over far more frequently with its perfective construction than with its imperfective one (see Table 4), which may hint at the inherent telicity of the verb. A certain number of nominal objects are distinctive of the PC-construction, some of which are related to media: *film*, *match*, *émission* ‘program’ and *reportage*. The only distinctive nominal collexeme of the IMP-construction is *gens* ‘people’, although other complements, of course, appear with a certain frequency. Finally, if the complementizer *que* ‘that’ is treated as a collexeme, the analysis reveals that it is attracted to the PC.

Table 4. Distinctive collexemes of the past-tense constructions of *voir* (Coll.Str. > 3)

| PC with nominal complementation (<i>N</i> = 14,453) | | IMP with nominal complementation (<i>N</i> = 1,793) | |
|------------------------------------------------------------|-----------|-------------------------------------------------------------|-----------|
| Collexeme | Coll.Str. | Collexeme | Coll.Str. |
| <i>jour</i> (‘day’) | 16.85 | <i>gens</i> (‘people’) | 3.80 |
| <i>film</i> (‘film’) | 14.39 | | |
| <i>match</i> (‘match’) | 10.53 | | |
| <i>émission</i> (‘show’) | 8.83 | | |
| <i>reportage</i> (‘report’) | 8.37 | | |
| <i>résultat</i> (‘result’) | 8.01 | | |
| <i>ventes</i> (‘sales’) | 6.73 | | |
| PC with verbal complementation (<i>N</i> = 4,794) | | IMP with verbal complementation (<i>N</i> = 807) | |
| <i>naître</i> (‘be born’) | 15.18 | <i>poindre</i> (‘break’) | 5.27 |
| <i>sortir</i> (‘come out’) | 5.44 | <i>briller</i> (‘shine’) | 3.78 |
| <i>jouer</i> (‘play’) | 5.28 | | |
| <i>entrer</i> (‘enter’) | 3.52 | | |
| <i>grandir</i> (‘grow up’) | 3.18 | | |
| PC with <i>que</i> -complementation (<i>N</i> = 6,512) | | IMP with <i>que</i> -complementation (<i>N</i> = 1,098) | |
| <i>que</i> (‘that’) | 8.51 | | |

The nouns related to media are all used in a PC-construction that construes the event as bounded as shown in (12). They are cases that can be explained aspectually and do not seem to have any phraseological characteristics.

- (12) *Ce n'est que hier que j'ai vu l'émission sur la tauromachie.* (Diaries)
‘It was only yesterday that I watched the show on bullfighting’.

The most distinctive complement *jour* ‘day’ forms the bounded phrasal expression *voir le jour* ‘emerge, come about’. The fact that the complement *ventes* ‘sales’ is also realized perfectly is initially counterintuitive, since one might expect the plural to trigger an imperfective reading. The greater picture, however, shows a pattern ‘company + *a vu ses ventes* + INF’ that is almost exclusive to business newspapers. This pattern is attested in the literature (Bat-Zeev Shyldkrot 1984: 205) and illustrates an example of *voir* as a grammaticalized auxiliary. Realized perfectly, it co-occurs with verbal complements like *augmenter* ‘increase’ or *baisser* ‘decrease’ that describe how a sales have changed, the event being construed as one undividable whole.

Once again, the verbal collexemes seem to be the most interesting ones: the perfective complements can be regrouped under the label of dynamic verbs that are relatively frequent. Their co-occurrence can be explained by the fact that the PC prototypically encodes the result of a past action. By contrast, the two IMP-collexemes *briller* ‘shine’ and *poindre* ‘break (day/dawn)’ are infrequent and evoke background descriptions in fiction. In fact, these patterns are typical examples of how a narrative temporarily zooms in, creating a period of stasis. Accordingly, they demonstrate how the French IMP is typically considered as the present of the past. Some instances of these past-tensed *voir*-constructions are listed in the concordance in Figure 2.

| | | | | |
|------------|----------------------------------|--------------|--------|-----------------------------------------|
| SMS | bien, parlé !! ;-) | Mdr j’ | ai vu | un reportage sur m 6 sur le |
| SMS | Ton numéro) | Fnalement tu | as vu | ton film ? C’est super si ton |
| Diaries | Ce n’est que hier que j’ | | ai vu | l’émission sur la tauromachie. |
| Magazines | Volkswagen depuis 1991, | | a vu | ses ventes augmenter de près |
| Newspapers | Décathlon et Intersport | | a vu | ses ventes et sa rentabilité |
| Newspapers | : PSA Peugeot Citroën | | a vu | ses ventes s’effondrer de 29,2 |
| Newspapers | Ce spécialiste de la pomme | | a vu | le jour à Neuve-église en 1989 |
| TV | Dans l’Empire byzantin | | ont vu | le jour en Languedoc. Deux |
| TV | ainsi que le bon géant | | a vu | le jour. Nous vous présentons |
| Formal | d’une technologie qui | | a vu | naître des aventures |
| TV | suite. Depuis 3 ans, on | | a vu | naître des ovnis dans le |
| Film | jetons-le à la poubelle. Je l’ | | ai vu | naître ce petit. Pistolet Jo, ne |
| Fiction | Abel battre dans sa poitrine. Il | | voyait | briller le petit nuage de vapeur devant |
| Fiction | . Sous la casquette, Bill | | voyait | briller des yeux bridés dans une face |
| Fiction | vers son malheur. Au loin, il | | voyait | briller les cygnes en groupe qui |
| Fiction | dans l’oreille d’un chat. Je la | | voyais | briller hors de ma portée, m’appeler |
| Fiction | citadelles de sable. Au loin, je | | voyais | briller les eaux de la mer Morte, qui |
| Fiction | péril mortel était écarté, Diane | | voyait | poindre en elle de nouveaux |
| Fiction | le disque de ciel bleu qu’il | | voyait | poindre au-dessus de lui. Une |
| Fiction | . En vain. Alors qu’il | | voyait | poindre la lisière de la forêt, il |
| Fiction | sur les fonds des mers: on | | voyait | poindre le moment où l’on pourrait |
| Fiction | En dépit de la barbarie qu’il | | voyait | poindre chez certains, toute son |

Figure 2. Concordance of the past-tense constructions of *voir*

Turning to clausal *que*-complementation (i.e. all the instances with restrictive *ne V que* 'only' were excluded), a conspicuous pattern can be noticed. Looking on the left side of the PC-construction at the preferred subject pronouns, we observe a clear preference for *nous / on* 'we'. This pattern *nous avons vu / on a vu que* 'we have seen that' has a text-structuring function, picking up a previously discussed topic and guiding the reader towards the next one (see Figure 3). Its relative text type frequency, as calculated with Sketch Engine, is more than 12 times higher in academic papers than in the corpus average.

| | | | |
|------|------------------------------------|----------|--------------------------------|
| Acad | nous est doué en quelque chose. On | a vu | que l'innovation principale |
| Acad | pour tous types de produits. On | a vu | que la commande de cassettes |
| Acad | assurée par le cliffement. On | a vu | que cette question n'était pas |
| Acad | deux ans plus tard. De plus, on | a vu | que la dynamique passée des |
| Acad | concerne l'Antiquité tardive, nous | avons vu | que la reprise des mythes |
| Acad | des autres participants. Nous | avons vu | que Samantha a affirmé son |
| Acad | et extratextuels de ce roman. Nous | avons vu | que les voix des personnages |

Figure 3. Concordance of the PC-construction of *voir* with clausal *que*-complementation

Que-complementation with the IMP occurs six times less frequently. These instances do not express literal vision of a physical object, but rather the speaker's capability of understanding a particular situation (see Figure 4). There is a close link between the perceptual and the cognitive interpretation, as in the third and fourth example of the concordance: the speaker observes a certain physical trait associated with a person and deduces that this person must have had a particular mood.

| | | | |
|--------|------------------------------------|--------|------------------------------------|
| Forums | ... meme face a Melenchon, on | voyait | qu'a part la lecture de ses |
| Forums | il prenait un bib le soir, mais on | voyait | qu'il se jetais pas dessus et puis |
| Forums | devait avoir 40 ans). Pourtant on | voyait | qu'elle était pas en bonne santé, |
| TV | Osterrieth : Outre le fait qu'on | voyait | qu'il était très tendu et qu'il |
| TV | - J'étais un peu timide, je | voyais | qu'ils étaient en train de me |
| TV | suis posé des questions car je | voyais | qu'il changeait. Un jour, je |

Figure 4. Concordance of the IMP-construction of *voir* with clausal *que*-complementation

These few examples indicate that the past-tense choice of a polysemous verb like *voir* correlates with certain lexico-grammatical constructions. The perfective constructions denote vision in the past as a bounded event, be it in its literal sense without any phraseological tendencies (*j'ai vu le film*) or with bounded phrasal expressions (*il a vu le jour*). Furthermore, it is commonly used as a rhetorical routine formula in academic writing (*nous avons vu que*). The imperfective construction occurs less frequently, attracts far less collexemes and illustrates a case of a common imperfective value.

5. French past-tensed verbs as lexico-grammatical constructions

What can the corpus-driven findings contribute to the analysis of French verbs and their past-tense constructions? Generally speaking, they can be *a posteriori* considered genuine constructions. Several pieces of lexico-grammatical evidence can be put forward in favour of this claim. First and foremost, it should have become clear that a linguistic phenomenon traditionally considered grammatical can be fruitfully analyzed lexico-grammatically. This is in keeping with one of the central tenets of CG, that is, the inseparability of lexis and grammar.

We can address the lexical grammar of past-tensed verbs from two sides, be it as a bottom-up *lexis-to-grammar* or, vice versa, as a top-down *grammar-to-lexis* approach (Hunston 2015). In the bottom-up perspective, a verb shows a tendency for one of the past tenses. In the top-down perspective, these patterns can be analyzed as collocations that co-occur preferably with a limited set of collocations.

Second, highly distinctive collexemes hint at patterns that are highly frequent and thus prone to be stored and entrenched separately as relatively fixed combinations, such as *je voulais vous demander si* or *nous avons vu que*. The usage-based criterion, according to the more recent definition given by Goldberg (2006), is thus satisfied.

Third, the more rigorous criterion of non-compositional semantics is fulfilled, although this point must be discussed in more detail: of course, not all past-tensed verbs display unpredictable semantics (cf. the example *j'ai vu le film* 'I saw the movie'). Yet there are instantiations, namely the highly frequent phraseological ones, which can best be located on a continuum of idiomaticity (Kay & Michaelis 2012: 2275–2277). At the one end of the continuum, past-tense constructions can be fully predictable, simply construing an event globally or partially; traditional aspectual approaches can solidly explain past-tense use in these cases.

On the other end, phraseological past-tense constructions can display usage constraints or unpredictable semantics. Consider the case studies discussed here: the fact that imperfective *vouloir* can encode politeness is not derivable from the past-tense construction itself. It cannot even be predicted, that, as the collocational analysis has shown, this meaning seems to be the most strongly entrenched one. Usage-based constraints, which in CG are also part of the constructional meaning component, can be identified in patterns such as 'company + *a vu ses ventes* + INF' exclusive to economic language, in the aforementioned polite imperfect with *vouloir* used in conversation, or in the argumentative pattern *nous avons vu que*.

In any case, this approach can help to disambiguate the past-tense semantics of highly frequent polysemous verbs, for example the alternation between *j'ai vu que* and *je voyais que*, as these patterns can not only be distinguished by their boundedness feature, but they also correlate with respective usage frames.

But what about infrequent verbs like, for instance, *décentraliser* ‘decentralize’? It is probably fair to say that phraseological patterns will be few or not even attested at all. Past-tense choice in such a case can best be explained by the traditional linguistic hypotheses on French PC and IMP, be it in the framework of discourse grounding or of concepts like the virtuality feature of the IMP. A careful hypothesis may therefore be put forward: the more frequent a verb, the more likely it is to be polysemous and the more likely it will have phraseological patterns that correlate with past-tense choice.

6. Concluding remarks

This chapter has demonstrated how a corpus-driven lexico-grammatical approach to French past-tensed verbs can help explore their usage patterns. Every verb can be realized in both the PC and the IMP, but speakers do not do so arbitrarily; instead, they resort to more or less schematized lexico-grammatical patterns. They can be realized with compositional semantics, but also as frozen phraseological units or even as genuine constructions with (usage-based) constraints.

From a methodological point of view, it is clear that our analysis has only scratched the surface of a vast area and offers one possible way towards a constructional view of past tenses. Advanced quantitative corpus methods can help to further refine the picture because they are able to consider additional variables such as mode or register in recent developments like multidimensional collocation analysis (Stefanowitsch & Gries 2008). In any case, it is hoped that the potential of collocation analysis could be pointed out. The analysis put forward in this chapter extended its use to non-verbal collexemes, that is, object noun phrases and, to a lesser extent, the complementizer *que* ‘that’, showing that collocation analysis can also be used for the analysis of phenomena that are not seen as pure syntactic alternations. In any case, it allows to identify phraseological usage patterns on statistically reliable grounds.

Finally, it should be mentioned that our approach may facilitate the acquisition of past tenses in instructed SLA. In fact, past-tense choice is one of the most daunting phenomena for all those learners whose L1 has a differently structured tense-aspect system than those shared by the Romance languages (cf. Blyth 2005 and Ayoun 2013 for an overview of core issues). One major challenge is the question which tense to use in order to describe some particular situation. If, however, the choice is not entirely dependent on the internal constituency of an event, but also predetermined by prototypical lexico-grammatical contexts, then the learner is well advised to learn these alternations with the semantic-pragmatic nuances instantiated by the construction in question. Of course, this new approach would

involve the revision of grammar rules as well as of pedagogical treatment principles in the classroom. A phraseodidactic approach seems to be particularly promising in this regard.

To conclude, it is hoped that a constructional approach can improve our understanding of past tenses, adding another piece to the puzzle of the lexical grammar of verbal constructions. Future studies may continue to venture into the interaction between past-tense choice and complementation pattern. One interesting corpus linguistic question would be to investigate what mechanisms coerce past-tense choice even if a given construction clearly co-occurs with one past tense, e.g. occurrences of *il voyait le jour* instead of prototypical *il a vu le jour*.

Acknowledgements

Many thanks go to Dalila Ayoun and the anonymous reviewers whose comments proved invaluable help to improve the paper considerably. All remaining inconsistencies and inadequacies are of course my own.

References

- Abouda, Lotfi. 2004. Deux types d'imparfait atténuatif. *Langue Française* 142: 58–74. doi:10.3917/lf.142.0058
- Anscombe, Jean-Claude. 2004. L'imparfait d'atténuation: Quand parler à l'imparfait, c'est faire. *Langue Française* 142: 75–99. doi:10.3917/lf.142.0075
- Antoniou, Maria. 2005. Le passé composé du français: Une forme verbale à valeur temporelle? In *Temporalité et attitude*, Arie Molendijk & Co Vet (eds), 171–185. Amsterdam: Rodopi.
- ATILF – CNRS & Univ. de Lorraine. 2016. *Base textuelle FRANTEXT*.
- Ayoun, Dalila. 2013. *The Second Language Acquisition of French Tense, Aspect, Mood and Modality* [AILA Applied Linguistics Series 10]. Amsterdam: John Benjamins. doi:10.1075/aals.10
- Bat-Zeev Shyldkrot, Hava. 1984. La concurrence entre la proposition conjonctive et *voir* + la proposition infinitive. *The French Review* 58(2): 202–214.
- Berthonneau, Anne-Marie & Kleiber, Georges. 1994. Imparfais de politesse: Rupture ou cohésion? *Travaux de Linguistique* 29: 59–92.
- Berthonneau, Anne-Marie & Kleiber, Georges. 2006. Sur l'imparfait contrefactuel. *Travaux de Linguistique* 53: 7–65. doi:10.3917/tl.053.0007
- Bilger, Mireille & Cappeau, Paul. 2013. Comment les données de corpus pourraient renouveler les manuels de grammaire? *Linx* 68–69: 177–199.
- Binnick, Robert I. (ed.). 2012. *The Oxford Handbook of Tense and Aspect*. Oxford: OUP. doi:10.1093/oxfordhb/9780195381979.001.0001
- Blumenthal, Peter. 1976. Imperfekt und Perfekt der französischen Modalverben. *Zeitschrift für französische Sprache und Literatur* 86(1): 26–39.

- Blyth, Carl. 2005. From empirical findings to the teaching of aspectual distinctions. In *Tense and Aspect in Romance Languages* [Studies in Bilingualism 29], Dalila Ayoun & Rafael Salaberry (eds), 211–252. Amsterdam: John Benjamins. doi:10.1075/sibil.29.08bly
- Boas, Hans C. & González-García, Francisco (eds). 2014. *Romance Perspectives on Construction Grammar* [Constructional Approaches to Language 15]. Amsterdam: John Benjamins.
- Boogaart, Ronny. 2004. Aspect and Aktionsart. In *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung*, Geert Booij & Herbert Wiegand (eds), 1165–1180. Berlin: Mouton Mouton de Gruyter.
- Bouveret, Myriam & Legallois, Dominique (eds). 2012. *Constructions in French* [Constructional Approaches to Language 13]. Amsterdam: John Benjamins. doi:10.1075/cal.13
- Broccias, Cristiano. 2012. The syntax-lexicon continuum. In *The Oxford Handbook of the History of English*, Terttu Nevalainen & Elizabeth Closs Traugott (eds), 735–747. Oxford: OUP.
- Bybee, Joan. 2006. From usage to grammar. The mind's response to repetition. *Language* 82: 711–733. doi:10.1353/lan.2006.0186
- Carlson, Lauri. 1984. Aspect and quantification. In *Syntax and Semantics*, Vol. 14, Philip J. Tedeschi & Annie Zaenen (eds), 31–65. New York NY: Academic Press.
- Chocheyras, Jacques. 1968. Un nouvel outil grammatical en français moderne: Le verbe *voir*. *Le Français Moderne* 36: 219–225.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: CUP.
- Croft, William & Cruse, Alan D. 2004. *Cognitive Linguistics*. Cambridge: CUP. doi:10.1017/CBO9780511803864
- Desclés, Jean-Pierre & Guentchéva, Zlatka. 2003. Comment déterminer les significations du passé composé par une exploration contextuelle? *Langue Française* 138: 48–60. doi:10.3406/lfr.2003.6481
- Detges, Ulrich. 2010. Computed or entrenched? The French *imparfait de politesse*. In *Cognitive Foundations of Linguistic Usage Patterns*, Hans-Jürgen Schmid & Susanne Handl (eds), 195–224. Berlin: Mouton de Gruyter. doi:10.1515/9783110216035.195
- De Knop, Sabine, Mollica, Fabio & Kuhn, Julia (eds). 2013. *Konstruktionsgrammatik in den romanischen Sprachen*. Frankfurt: Peter Lang.
- Deulofeu, Henri-José & Debaisieux, Jeanne-Marie. 2012. Une tâche à accomplir pour la linguistique française du XXI^e siècle: Élaborer une grammaire des usages du français. *Langue Française* 176: 27–46.
- Di Vito, Nadine O'Connor. 1997. *Patterns Across Spoken and Written French: Empirical Research on the Interaction among Forms, Functions, and Genres*. New York NY: Houghton Mifflin.
- Diessel, Holger. 2015. Usage-based construction grammar. In *Handbook of Cognitive Linguistics*, Eva Dąbrowska & Dagmar Divjak (eds), 296–322. Berlin: Mouton de Gruyter. doi:10.1515/9783110292022-015
- Do-Hurinville, Danh-Thành. 2010. Etude des temps verbaux dans les articles nécrologiques. *Syntaxe & Sémantique* 11: 83–111. doi:10.3917/ss.011.0083
- Ellis, Nick, Römer, Ute & O'Donnell, Matthew Brook. 2016. *Usage-Based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. Hoboken NJ: Wiley & Sons.
- Eshkol-Taravella, Iris, Baude, Olivier, Maurel, Denis, Hriba, Linda, Dugua, Céline & Tellier, Isabelle. 2011. Un grand corpus oral “disponible”: Le corpus d'Orléans 1968–2012. *TAL* 53(2): 17–46.

- Facques, Bénédicte. 2002. Passé composé, imparfait et présent dans les récits journalistiques: Des alternances aux ruptures temporelles. In *Les temps du passé français et leur enseignement*, Emmanuelle Labeau & Pierre Larrivé (eds), 105–133. Amsterdam: Rodopi.
- Falaise, Achille. 2005. Constitution d'un corpus de français tchaté. *RÉCITAL 2005*, Dourdan, France.
- Fillmore, Charles, Kay, Paul & O'Connor, Mary Catherine. 1988. Regularity and idiomaticity in grammatical constructions: The case of *Let Alone*. *Language* 64(3): 501–538. doi:10.2307/414531
- Fleischman, Suzanne. 1985. Discourse functions of tense-aspect oppositions in narrative: Toward a theory of grounding. *Linguistics* 23(6): 851–882. doi:10.1515/ling.1985.23.6.851
- Garey, Howard B. 1957. Verbal aspect in French. *Language* 33(2): 91–110. doi:10.2307/410722
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago IL: Chicago University Press.
- Goldberg, Adele. 2006. *Constructions at Work*. Oxford: OUP.
- Gries, Stefan T. 2007. *Coll.Analysis 3.2a. A program for R for Windows 2.x*.
- Gries, Stefan T. 2008. Phraseology and linguistic theory. A brief survey. In *Phraseology: An Interdisciplinary Perspective*, Sylviane Granger & Fanny Meunier (eds), 3–25. Amsterdam: John Benjamins. doi:10.1075/z.139.06gri
- Gries, Stefan T. 2010. Methodological skills in corpus linguistics: A polemic and some pointers towards quantitative methods. In *Corpus Linguistics in Language Teaching*, Tony Harris & María Moreno Jaén (eds), 121–146. Frankfurt: Peter Lang.
- Gries, Stefan T. & Stefanowitsch, Anatol 2004. Extending collocation analysis. A corpus-based perspective on alternations. *International Journal of Corpus Linguistics* 9(1): 97–129.
- Gross, Maurice. 1985. *Méthodes en syntaxe*. Paris: Klincksieck.
- Hare, Mary, McRae, Kevin & Elman, Jeffrey L. 2003. Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language* 48: 281–303. doi:10.1016/S0749-596X(02)00516-8
- Herzog, Christian. 1981. *Le passé simple dans les journaux du XXe siècle*. Bern: Franke.
- Hopper, Paul J. 1979. Aspect and foregrounding in discourse. In *Syntax and Semantics: Discourse and syntax*, Talmy Givón (ed.) 214–241. New York NY: Academic Press.
- Hunston, Susan. 2015. Lexical grammar. In *The Cambridge Handbook of English Corpus Linguistics*, Douglas Biber & Randi Reppen (eds), 201–215. Cambridge: CUP. doi:10.1017/CBO9781139764377.012
- Imbs, Paul. 1960. *Les emplois des temps verbaux en français moderne. Essai de grammaire descriptive*. Paris: Klincksieck.
- Kamp, Hans & Rohrer, Christian. 1983. Tense in Texts. In *Meaning, Use, and Interpretation of Language*, Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds), 250–269. Berlin: Mouton de Gruyter. doi:10.1515/9783110852820.250
- Kay, Paul & Fillmore, Charles. 1999. Grammatical constructions and linguistic generalizations: The 'What's X Doing Y'? construction. *Language* 75(1): 1–33. doi:10.2307/417472
- Kay, Paul & Michaelis, Laura A. 2012. Constructional meaning and compositionality. In *Semantics, Vol. 3*, Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds), 2271–2296. Berlin: Mouton de Gruyter.
- Klum, Arne. 1961. *Verbe et adverbe*. Stockholm: Almqvist & Wiksell.
- Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. In *Semantics and Contextual Expression*, Renate Bartsch, Johan van Benthem & Peter van Emde Boas (eds), 75–115. Dordrecht: Foris.

- Labeau, Emmanuelle. 2002. L'unité de l'imparfait: Vues théoriques et perspectives pour les apprenants du français langue étrangère. *Travaux de Linguistique* 45(2): 157–184. doi:10.3917/tl.045-157
- Labeau, Emmanuelle. 2004. Les temps du compte rendu sportif francophone. *Journal of French Language Studies* 14(2): 129–148. doi:10.1017/S0959269504001619
- Labeau, Emmanuelle. 2005. Mon nom est narratif: Imparfait narratif. In *Nouveaux développements de l'imparfait*, Emmanuelle Labeau & Pierre Larrivée (eds), 79–102. Amsterdam: Rodopi.
- Labeau, Emmanuelle. 2006. French television talk: What tenses for past time? *International Journal of Corpus Linguistics* 11(1): 1–28. doi:10.1075/ijcl.11.1.02lab
- Labeau, Emmanuelle. 2007. Et un, ou deux, ou trois? Les temps-champions du compte rendu sportif depuis 1950. In *Sémantique et diachronie du système verbal français*, Emmanuelle Labeau, Carl Vetters & Patrick Caudal (eds), 203–221. Amsterdam: Rodopi.
- Labeau, Emmanuelle. 2009. Le PS: cher disparu de la rubrique nécrologique? *Journal of French Language Studies* 19(1): 61–86. doi:10.1017/S0959269508003578
- Labeau, Emmanuelle. 2013. Regrets éternels: le PS disparaît-il de la rubrique nécrologique? In *Marqueurs temporels et modaux en usage*, Walter de Mulder, Jesse Mortelmans & Tanja Mortelmans (eds), 95–118. Amsterdam: Rodopi.
- Labelle, Marie. 1987. L'utilisation des temps du passé dans les narrations françaises: Le passé composé, l'imparfait et le présent historique. *Revue Romane* 22(1): 3–29.
- Labelle, Marie. 1996. Remarques sur les verbes de perception et la sous-catégorisation. *Recherches linguistiques de Vincennes* 25: 83–106.
- Larrivée, Pierre. 2002. Sémantique conceptuelle et sémantique référentielle du passé composé. In *Les temps du passé français et leur enseignement*, Emmanuelle Labeau & Pierre Larrivée (eds), 51–69. Amsterdam: Rodopi.
- Legallois, Dominique & Patard, Adeline (eds). 2017. Les constructions comme unités de la langue: Illustrations, évaluation, critique. *Special issue of Langue Française* 194.
- Levshina, Natalia. 2017. Subtitles as a corpus: An n-gram approach. *Corpora* 12(3): 311–338.
- Luscher, Jean-Marc. 1998. Procédure d'interprétation du passé composé. In *Le temps des événements. Pragmatique de la référence temporelle*, Jacques Moeschler (ed.), 181–196. Paris: Kimé.
- Martinez, Ron & Schmitt, Norbert. 2012. A phrasal expressions list. *Applied Linguistics* 33(3): 299–320. doi:10.1093/applin/ams010
- Michaelis, Laura A. 1998. *Aspectual Grammar and Past-Time Reference*. London: Routledge.
- Molendijk, Arie. 2002. La structuration logico-temporelle du texte: Le passé simple et l'imparfait du français. In *Les temps du passé français et leur enseignement*, Emmanuelle Labeau & Pierre Larrivée (eds), 91–104. Amsterdam: Rodopi.
- New, Boris, Brysbaert, Marc, Veronis, Jean & Pallier, Christoph. 2007. The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28(4): 661–677. doi:10.1017/S014271640707035X
- Panckhurst, Rachel, Détrie, Catherine, Lopez, Cédric, Moïse, Claudine, Roche, Mathieu & Verine, Bertrand. 2013. Sud4science, de l'acquisition d'un grand corpus de SMS en français à l'écriture SMS. *Épistémè – Revue Internationale de Sciences Sociales Appliquées* 9: 107–138.
- Patard, Adeline. 2007. *L'un et le multiple. L'imparfait de l'indicatif en français: Valeur en langue et usages en discours*. PhD dissertation, Université Paul-Valéry-Montpellier III.
- Patard, Adeline. 2010. L'emploi préliminaire de l'imparfait entre temporalité et modalité: Éléments d'analyse à partir d'une étude de cas. *Journal of French Language Studies* 20(2): 189–211. doi:10.1017/S0959269509990494

- Quaglio, Paulo. 2009. *Television Dialogue: The Sitcom Friends vs. Natural Conversation* [Studies in Corpus Linguistics 36]. Amsterdam: John Benjamins. doi:10.1075/scl.36
- Rebotier, Aude. 2005. *Les temps verbaux en allemand et en français et les phénomènes aspectuels*. PhD dissertation, Université Grenoble III.
- Rice, Sally & Newman, John. 2005. Inflectional islands. Paper presented at the International Cognitive Linguistics Conference, Seoul, South Korea.
- Römer, Ute. 2009. The inseparability of lexis and grammar. Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics* 7: 140–162. doi:10.1075/arcl.7.06rom
- Sasse, Hans-Jürgen. 2002. Recent activity in the theory of aspect: Accomplishments, achievements, or just non-progressive state? *Linguistic Typology* 6: 199–271. doi:10.1515/lity.2002.007
- Siepmann, Dirk. 2015. Dictionaries and spoken language: A corpus-based review of French dictionaries. *International Journal of Lexicography* 28(2): 139–168. doi:10.1093/ijl/ecvoo6
- Siepmann, Dirk & Bürgel, Christoph. 2015. L'élaboration d'une grammaire pédagogique à partir de corpus: l'Exemple du subjonctif. In *Grammatikographie und Didaktische Grammatik – gestern, heute, morgen. Gedenkschrift für Hartmut Kleineidam anlässlich seines 75. Geburtstages*, Thomas Tinnefeld (ed.), 159–185. Saarbrücken: htw saar.
- Siepmann, Dirk & Bürgel, Christoph. 2016. Das *Corpus de référence du français contemporain* (CRFC) und sein Einsatz in der Grammatikographie am Beispiel des Präpositionsgebrauchs. In *Sprachwissenschaft und Fremdsprachendidaktik. Zum Verhältnis von sprachlichen Mitteln und Kompetenzentwicklung*, Dirk Siepmann & Christoph Bürgel (eds), 143–162. Baltmannsweiler: Schneider Hohengehren.
- Siepmann, Dirk & Bürgel, Christoph. To Appear. Les unités phraséologiques fondamentales du français contemporain. In *Approches théoriques et empiriques en phraséologie*, Maurice Kauffer & Yves Keromnes (eds). Tübingen: Stauffenberg.
- Siepmann, Dirk, Bürgel, Christoph & Diwersy, Sascha. 2017. The corpus de référence du français contemporain (CRFC) as the first genre-diverse mega-corpus of French. *International Journal of Lexicography* 30(1): 63–84.
- Smith, Carlota S. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer. doi:10.1007/978-94-015-7911-7
- Stefanowitsch, Anatol & Gries, Stefan T. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2): 209–243. doi:10.1075/ijcl.8.2.03ste
- Stefanowitsch, Anatol & Gries, Stefan T. 2008. Channel and constructional meaning: A collostructional case study. In *Cognitive Sociolinguistics*, Gitte Kristiansen & René Dirven (eds), 129–152. Berlin: Mouton de Gruyter. doi:10.1515/9783110199154.2.129
- Stein, Achim. 2003. *French TreeTagger Part-of-Speech Tags, version of April 2003*. <www.ims.uni-stuttgart.de/~schmid/french-tagset.html>
- Sten, Holger. 1952. *Les temps du verbe fini (indicatif) en français moderne*. København: Munksgaard.
- Stubbs, Michael. 2015. Corpus semantics. In *The Routledge Handbook of Semantics*, Nick Riemer (ed.), 106–121. London: Routledge.
- Taylor, John R. 2012. *The Mental Corpus. How Language is Represented in the Mind*. Oxford: OUP. doi:10.1093/acprof:oso/9780199290802.001.0001
- de Saussure, Louis. 2000. Quand le temps ne progresse pas avec le passé simple. In *Passé et parfait*, Anne Carlier, Véronique Lagae & Céline Benninger (eds), 37–48. Amsterdam: Rodopi.
- de Swart, Henriette. 1998. Aspect Shift and Coercion. *Natural Language & Linguistic Theory* 16(2): 347–385. doi:10.1023/A:1005916004600
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work* [Studies in Corpus Linguistics 6]. Amsterdam: John Benjamins. doi:10.1075/scl.6

- Vet, Co. 1992. Le passé composé: Contextes d'emploi et interprétation. *Cahiers de Praxématique* 19: 37–59.
- Vetters, Carl. 1996. *Temps, aspect et narration*. Amsterdam: Rodopi.
- Vetters, Carl & de Mulder, Walter. 2000. Passé simple et imparfait: Contenus conceptuel et procédural. In *Passé et parfait*, Anne Carlier, Véronique Lagae & Céline Benninger (eds), 13–36. Amsterdam: Rodopi.
- Waugh, Linda R. & Monville-Burston, Monique. 1986. Aspect and discourse function: The French simple past in newspaper usage. *Language* 62(4): 846–877. doi:10.2307/415174
- Weinrich, Harald. 1964. *Tempus: Besprochene und erzählte Welt*. Stuttgart: Kohlhammer.
- Willems, Dominique. 2012. Verb typology. Between construction and lexicon. In Bouveret & Legallois (eds), 23–48.
- Wilmet, Marc. 1997. *Grammaire critique du français*. Louvain-la-Neuve: Duculot.

Path scales

Directed-motion verbs, prepositions and telicity in European Portuguese

António Leal, Fátima Oliveira and Purificação Silvano
University of Porto

We put forward a proposal based on the notion of path scale (e.g., Kennedy & Levin 2008) to account for the semantics of inherently directed motion verbs *ir* ('go') and *vir* ('come') combined with prepositional phrases (PPs) with the thematic role of Goal, headed by the prepositions *para* ('to/toward') and *até* ('to') in European Portuguese. We describe our corpus before discussing the syntactic and semantic properties of the predications we analyze as well as their similarity to predications with manner of motion verbs. We assume that motion verbs have an underspecified underlying scalar structure and that PPs contribute in a distinct way to the specification of one of the parameters of these path scales.

Keywords: inherently directed-motion verbs, scale semantics, prepositions, telicity

1. Introduction^{1,2}

Identifying the contribution of prepositions to the determination of the aspectual profile of predications³ that denote motion events poses certain theoretical problems. The main problem concerns the telicity of predications, that is, although there are prepositions that introduce prepositional phrases (PPs) with the thematic role of Goal, they do not always determine the predication's telicity because, in some cases,

-
1. We are grateful to three anonymous referees and the editors for very helpful comments on the material discussed in this chapter.
 2. This chapter is based upon work carried out within project 139614 supported by Fundação Calouste Gulbenkian.
 3. We use the term 'predication' to refer to the situation denoted by the combination of the NP subject and the VP (verb + complements).

the telic interpretation depends on contextual factors. So the issue is to explain the different behavior of these prepositions concerning telicity of the predications.

To this end, we analyze verbs of inherently directed motion (Levin 1993) in European Portuguese (EP), specifically *ir* ('go') and *vir* ('come'), that exhibit a complex pattern of possibilities of combination when they occur with PPs with the thematic role of Goal, in particular when headed by the prepositions *para* ('to/toward') and *até* ('to'). These possibilities of combination with *para* and *até* can also be found in predications with manner of motion verbs (cf. Leal & Oliveira 2015).

Our main objectives are: (a) to describe the interpretations that can arise from the combination of the verbs of movement *ir* and *vir* with PPs headed by *para* ('to/toward') and *até* ('to') (when denoting motion events); (b) to highlight some aspectual characteristics of these constructions; and (c) to put forward an explanation of the data grounded in scalar semantics.

We analyze two types of examples: (a) examples illustrating the general behavior of verbs *ir/vir* with prepositions *para/até* drawn from a corpus that we built during January 2016 with sentences extracted from an online newspaper <www.observador.pt>; these examples will be identified throughout the chapter with *corp*; (b) constructed examples, based on examples from the corpus, used to illustrate semantic properties of predications with *ir/vir*; for these examples, the acceptability judgments of EP's native speakers (both linguists and non-linguists) were used to verify those of the authors. Whenever speakers had a different interpretation, we signaled it.⁴

This chapter is structured as follows. We begin with a general description of predications with verbs *ir* and *vir* in EP. In the following section, we present a characterization of predications with *ir* and *vir* containing PPs headed by *para* and *até*, focusing on syntactic and semantic properties. In the fourth section, we compare verbs of inherently directed motion *ir* and *vir* with verbs of manner of motion in EP and, in the fifth section, we offer an account of the data. Concluding remarks focus on the scalar meaning of inherently directed motion verbs and on the manner in which prepositions complete underspecified parameters of scales.

2. General description of predications with verbs *ir* and *vir* in EP

The corpus includes 514 tokens with motion verbs *ir* and *vir* extracted from an online newspaper <www.observador.pt>. The data extracted are from news published during only January 2016. The corpus allowed us to verify the actual use of prepositions *até* and *para* with verbs *ir* and *vir* in EP. We use a tool specifically developed for

4. Although we did not carry out a quantitative analysis of the elicitation tasks, we can say that at least twenty native speakers validated the interpretations that we present here.

- (3) a. *No sábado o Sporting vai até Paços de Ferreira.* (corp)
 On Saturday the Sporting go-PRS.3SG to Paços de Ferreira.
 ‘On Saturday Sporting goes to Paços de Ferreira.’
- b. **No sábado o Sporting vai para Paços de Ferreira.*
 On Saturday the Sporting go-PRS.3SG to Paços de Ferreira.
 ‘On Saturday Sporting goes to Paços de Ferreira.’

In these cases, the events denoted are events of movement and *ir* and *vir* behave the same way. However, when the events denoted are not events of movement, these two motion verbs behave differently. In fact, the verb *ir* provides contexts in which the prepositions *para* and *até* are not interchangeable, as exemplified by (4) and (5).

- (4) *A pena pode ir até 15 dias de prisão.*
 The penalty may-PRS.3SG go-INF to 15 days of prison.
 ‘The penalty may be as high as 15 days in prison.’ (corp) (**ir para (go to)*)
- (5) *José Sá tinha acordo para ir para outro clube.*
 José Sá have-PST.IPFV.3SG agreement to go-INF to another club.
 ‘José Sá had an agreement to go to another club.’ (corp) (**ir até (go to)*)

When the verb is *vir*, only the preposition *para* can occur in predications that do not denote events of movement.

- (6) *Essa receita tem que vir para o orçamento de estado.* (corp) (cf. **essa receita tem que vir até ao orçamento de estado.*)
 This revenue must-PRS.3SG that come-INF to the budget of state.
 ‘This revenue must go to the state budget.’
 ‘This revenue must come to the state budget’

Having observed that the structures under scrutiny accept both prepositions when the events represented are events of movement, we will turn to the main question of this chapter: semantic differences between contexts like those in (7) and those in (8). In such contexts, the presence of the two prepositions leads to different interpretations: in (7), the boy may or may not reach the gym,⁷ whereas in (8) the only available reading is the one in which the boy reaches destination.

- (7) *O rapaz foi /veio para o ginásio*
 ‘The boy went/ came to/toward the gymnasium’
- (8) *O rapaz foi /veio até ao ginásio*
 ‘The boy went/ came to the gymnasium’

7. Whenever *para* has an atelic interpretation, we use ‘toward’ in the translation; in the telic interpretation, we use ‘to’.

3. Toward a syntactic and semantic characterization of predications containing *ir* and *vir* in combination with PPs headed by *para* and *até*

In this section, our goal is twofold: to identify syntactic and semantic features of predications with *ir* and *vir* containing PPs headed by *para* and *até* and, simultaneously, to pinpoint some problems that arise from the differences between these two types of predications.

3.1 Events of movement: Some syntactic issues

Although we are mainly concerned with semantic properties of predications in which *ir/ vir* combine with PPs headed by *até/ para*, we need to address two syntactic features of these structures, because they are relevant to their semantic characterization.

The first one concerns the syntactic function fulfilled by the PP: when PPs headed by *para* and *até* combine with these verbs, they are complements, as demonstrated by tests commonly used to diagnose the complement status of constituents, as shown in (9) and (10).

- (9) * *O João {foi/veio} para casa e a Maria fez o mesmo para a escola. (fazer o mesmo = {ir/vir})*
 The João{ go/come-PST.3SG} to home and the Maria do-PST.3SG the same to the school. (do the same = {come / go})
 ‘João {went/came} home and Maria did the same to school. (do the same = {come/ go})’

- (10) * *O que é que o João fez até casa? {Foi / veio}.*
 What be-PRS.3SG that the João do-PST.3SG to home? {Come / go}
 ‘What did João do (up) to home? {Come / Go}’

Sentence (9) is ungrammatical because the expression ‘fazer o mesmo’ cannot refer only to the verb (‘*ir/ vir*’), but must instead refer to the entire VP. In (10), the insertion of *até casa* in the question renders the reply ungrammatical.

The second feature has to do with the deictic nature of *ir* and *vir*, which allows them to occur, in certain contexts, without any complement. In the absence of a PP complement, the verb *ir* requires the insertion of the adverb *já*, ‘already’ (cf. 12).

- (11) *O rapaz {*foi / veio}.*
 The boy {*go / come-PST.3SG}.
 The boy {* went /came}.

The verb *vir* imposes no such constraint:

- (12) *O rapaz já {foi / veio}.*
 The boy already {go / come-PST.3SG}.
 ‘The boy has already {gone/come}’.

This difference results from the fact that the verb *vir* contains lexical information of movement toward the location of the speaker, whereas the verb *ir* lexicalizes information of movement away from the speaker. Therefore, with *vir*, in the absence of an overt goal PP, it is assumed that the goal corresponds to the speaker’s location; with *ir*, the information about the goal is not available.

Having established that both PPs headed by *para/ até* are complements of *ir/ vir*, we will address some semantic properties of these predications in the following section.

3.2 Events of movement: Some semantic issues

The semantic analysis of predications containing the verbs *ir* and *vir* followed by PPs headed by *para* and *até* raises some problems that we will discuss in this section. The first one concerns the aspectual classes of these predications. Tests proposed in the literature to identify aspectual classes (Vendler 1957; Dowty 1979, a.o.), which work well when applied to the majority of cases in EP, but present some inconclusive results when applied to predications in which the verbs *ir* and *vir* select *para/ até*.

Concerning the combination of the PP with *in x time* adverbials, what is measured is the whole event, which shows that they are accomplishments (i.e., durative and telic events).

- (13) *O rapaz foi/ veio até à faculdade em 5 m.*
 The boy go/come-PST.3SG to to-the campus in 5 min.
 ‘The boy went/came to campus in 5 min’.
- (14) ^{ok/?}*O rapaz foi/veio para a faculdade em 5m.⁸*
 The boy go/come-PST.3SG to the campus in 5 min.
^{ok/?}‘The boy went/came to campus in 5 min’.

However, this combination of *Pretérito Perfeito* (Past) with *in x time* adverbial triggers different entailments: only *para* (cf. (16)) entails the truth of the Progressive during the same time, which allows for the identification of accomplishments. We interpret sentence (15) using world knowledge as follows: the entity denoted by *the*

8. Some informants find this sentence a little odd. We will return to this reading in Section 5.2. to try to explain it.

boy moved along a path that ends at campus and that movement lasts five minutes, but we cannot entail that during those five minutes the boy was coming to campus.⁹

- (15) *O rapaz veio até à faculdade em 5 m. → O rapaz*
 The boy come-PST.3SG to to-the campus in 5 min. The boy
esteve a vir até à faculdade durante esses 5 m.
 be-PST.3SG to come-INF to to-the campus during those 5 min.
 ‘The boy came to campus in 5 min’.
 ‘The boy was coming to campus during those 5 min’.
- (16) ^{ok/?}*O rapaz veio para a faculdade em 5m. → O rapaz*
 The boy come-PST.3SG to the campus in 5 min. The boy
esteve a vir para a faculdade durante esses 5 m.
 be-PST.3SG to come-INF to the campus during those 5 min.
 ‘The boy came to campus in 5 min’.
 ‘The boy was coming to campus during those 5 min’.

In EP, the combination of the *Pretérito Perfeito* (Past) with *for x time* adverbials can have different interpretations according to the aspectual nature of the predication. With achievements as in (17), there is typically a reading of measuring the resultant state, which is an atelic, durative and homogeneous phase of any event that follows a culmination point (cf. Moens 1987).

- (17) *O rapaz desmaiou durante 5 m.*
 The boy faint-PST.3SG during 5 min.
 Lit. ‘The boy fainted for 5 min’.

On the other hand, in EP, this reading of measuring the resultant state does not occur with accomplishments, which, when combined with *for x time* adverbials, have irresultative readings as in (18).

- (18) *O rapaz leu o livro durante 5 m.*
 The boy read-PST.3SG the book during 5 min.
 ‘The boy read the book for 5 min’ (the boy did not finish reading the book)

The combination of the *Pretérito Perfeito* (Past) with *for x time* adverbials leads to different results with *ir/ vir, para/ até* and *for x time*: (a) with *PP-até*, as in (19), the examples are grammatical, but only with a resultant state reading (the adverbial measures the duration of the state resulting from the event) (cf. Dowty 1979); (b) with *PP-para*, as in (20), there is some ambiguity between a reading of measuring

9. This is a puzzling question for which we do not have a satisfactory answer at this moment.

the preparatory process and a reading of measuring the resultant state.¹⁰ This possibility of resultant state reading seems to indicate that the predications are non-durative.

(19) *O rapaz foi/veio até à faculdade*

The boy go/come-PST.3SG to to-the campus

durante 5 m. (= esteve lá 5 m)

during 5 min.

'The boy went/came to campus (only) for 5 min'. (= he was there 5 min.)

(20) *O rapaz foi/veio para a faculdade durante*

The boy go/come-PST.3SG toward/to the campus during

5 m. (= esteve a ir para a faculdade durante 5 m. / esteve lá 5m)

5 min.

'The boy went/came toward/to campus during/for 5 min'. (= he was going to campus during 5 min./ was there for 5 min.)

Furthermore, the possibility of a non-culminating reading (i.e., of canceling the event's culmination) varies according to the prepositions: it is possible with PP-*para*, but not with PP-*até*.

(21) *O rapaz veio/foi para a faculdade, mas,*

The boy come/go-PST.3SG toward the campus, but,

a meio do caminho, teve de voltar para trás.

middle of-the path, had of come to back.

'The boy came/went toward campus, but halfway he had to retrace his steps'

(22) * *O rapaz veio/foi até à faculdade, mas,*

The boy come/go-PST.3SG to to-the campus, but,

a meio do caminho, teve de voltar para trás.

middle of-the path, had of come to back.

'The boy came/went to campus, but halfway he had to retrace his steps'.

Regarding durativity, we have also observed that the combination with the aspectual operator *parar de* ('to stop V-ing') produces unacceptable or even ungrammatical results, as illustrated by examples (23) and (24), with a single-event reading for both prepositions. This is the typical behavior of non-durative events.

10. An anonymous reviewer suggested that the resultant state reading in the EP sentences (17) and (18) could have a modal component that could be paraphrased as he intended to stay in campus for 5 minutes (not that he actually did), similar to the English version of the examples. However, this is not the case in EP, since this modal reading is completely absent.

- (23) a. ^{???}/*Ele parou de ir para a faculdade.
He stop-PST.3SG of go-INF to the campus.
'He stopped going toward campus'.
- b. ^{???}/* Ele parou de vir para a faculdade.
He stop-PST.3SG of come-INF to the campus.
'He stopped coming toward campus'.
- (24) a. ^{???}/* Ele parou de ir até à faculdade.
He stop-PST.3SG of go-INF to to-the campus.
'He stopped going to campus'.
- b. ^{???}/* Ele parou de vir até à faculdade.
He stop-PST.3SG of come-INF to to-the campus.
'He stopped coming to campus'.

In (23) and (24), the combination of *ir/ vir* followed by *para/ até* with *parar de V-Inf* ('to stop V-ing') shows similar results to the ones we can find in predications that are clearly achievements as in *vencer a corrida* 'win the race' (**parar de vencer a corrida* 'stop winning the race').

Regarding the Progressive, the two prepositions produce different results, since only *para* can occur with the Progressive (cf. (25)) and *até* gives rise to almost ungrammatical examples with a single event reading (cf. (26)).¹¹

- (25) a. *Ele está a ir para a faculdade.*
He be-PRS.3SG to go-INF to the campus.
'He's going toward campus'.
- b. *Ele está a vir para a faculdade.*
He bePRS.3ps to come-INF to the campus.
'He's coming toward campus'.
- (26) a. ^{???}*Ele está a ir até à faculdade.*
He be-PRS.3SG to go-INF to to-the campus.
'He's going to campus'.
- b. ^{???}*Ele está a vir até à faculdade.*
He be-PRS.3SG to come-INF to to-the campus.
'He's coming to campus'.

Another problem is the different (preferential) interpretations with punctual adverbs and *vir*. The examples (27) show that, if the preposition used is *para*, the punctual adverbial can mark either the time of departure or the time of arrival

11. With frequency adverbials, all examples are grammatical (cf. (a)).

(a) *Ele parou de {ir/vir} até à faculdade todas as manhãs.*
He stopped {going/coming} to college every morning.

(27a), while *até* is compatible only with an arrival reading (27b). By contrast, as shown in (28), if the verb is *ir*, a punctual adverbial receives the same (departure) reading regardless of the preposition used.

- (27) a. *O rapaz veio para a faculdade às 2h.*
 The boy come-PST.3SG to the campus at-the 2 o'clock.
 – *partida (ou chegada)*
 – departure (or arrival)
 ‘The boy came to campus at 2 o'clock’.
- b. *O rapaz veio até à faculdade às 2h.*
 The boy come-PST.3SG to to-the campus at-the 2 o'clock.
 – *chegada*
 – arrival
 ‘The boy came to campus at 2 o'clock’.
- (28) a. *O rapaz foi para a faculdade às 2h.*
 The boy go-PST.3SG to the campus at-the 2 o'clock.
 – *partida*
 – departure
 ‘The boy went to campus at 2 o'clock’.
- b. *O rapaz foi até à faculdade às 2h.*
 The boy go-PST.3SG to to-the campus at-the 2 o'clock.
 – *partida*
 – departure
 ‘The boy went to campus at 2 o'clock’.

Our final observation involves the PP’s specifying only direction or orientation. In order to achieve this reading, *ir/ vir* combine with *para* as illustrated in (29), but this interpretation is never possible with *até*, as shown in (30).

- (29) a. *O rapaz foi para oeste*
 The boy go-PST.3SG toward west
 ‘The boy went toward west’
- b. *O rapaz veio para oeste*
 The boy come-PST.3SG toward west
 ‘The boy came toward west’
- (30) a. **O rapaz foi até oeste*
 The boy go-PST.3SG to west
 ‘The boy went west’
- b. **O rapaz veio até oeste*
 The boy come-PST.3SG to west
 ‘The boy came west’

To sum up, the analysis of the data indicates the following problems. On one hand, syntactic criteria suggest that *para* and *até*, when combined with verbs *ir/vir*, behave identically, as complements of these verbs. On the other hand, semantic criteria point to distinct aspectual contributions: (a) only *para* can occur with a mere directional interpretation; (b) only *para* can occur with the Progressive; (c) only *para* gives rise to non-culminating readings. Furthermore, tests commonly used to diagnose aspectual properties of predications give rise to contradictory results. The *in x time* adverbial diagnostic suggests that all events are durative and telic. The *for x time* adverbial diagnostic suggests that event predications with *para* are ambiguous, allowing both a durative and telic reading and a non-durative and telic reading. By contrast, when the *for x time* adverbial is combined with *até*, event predications are unambiguously interpreted as non-durative and telic. When they are combined with *parar de*, both verbs with *para* and *até* give rise to ungrammatical sentences, which is typical of non-durative predications. Finally, although the *in x time* adverbial diagnostic suggests that all events are durative and telic, the analysis of the entailments with these adverbials when combined with the *Pretérito Perfeito* (Past) leads to different results: with *para*, the event is interpreted as durative and telic, but with *até*, it is non-durative and telic.

Overall, we have observed that, although predications with *ir/ vir* followed by PPs headed by *para/ até* share some features, they carry some relevant semantic differences. Predications with *para* denote events that can have or not have a terminal point. Predications with *até* denote events that have a terminal point and do not allow the terminal point to be omitted.

The data that we analyzed indicate that predications with *ir* and *vir* combined with *até* are accomplishments. Example (1) (*Ela foi /veio até Paris*), without time adverbials, is interpreted as a single durative event in which the entity that undergoes movement reaches destination. However, when combined with *for x time* adverbial, it seems that there is some kind of aspectual shift. Following Moens's (1987) proposal of aspectual nucleus, we can consider the *for x time* adverbial an operator that takes as input an accomplishment yielding as output the consequent state of the aspectual nucleus of the accomplishment. As for *ir* and *vir* combined with *para* (*Ela foi /veio para Paris*), predications seem to behave like activities that can be commuted into accomplishments when they co-occur with *in x time* adverbials because a culmination point is added to the aspectual nucleus by the adverbial.

4. Comparison between verbs of inherently directed motion *ir* and *vir* and verbs of manner of motion in EP

The verbs under analysis exhibit some similarities to manner of motion verbs in EP. However, they also show some differences (cf. Leal & Oliveira 2008, 2015). In this section, we will consider some of these similarities and distinctions in order to clarify some semantic issues concerning the verbs of inherently directed motion that we analysed.

- (31) * *O que é. fez para a faculdade? /*
 What be-PRS.3ps that the João do-PST.3SG to the campus? /
Caminhou.

Walked-PST.3SG

‘* What did João do to campus? {he walked}’

- (32) *O João caminhou para a faculdade, e a Maria fez*
 The João walk-PST.3SG to the campus, and the Maria do-PST.3SG
o mesmo para o café. (fazer o mesmo = caminhar)
 the same to the café. (do the same = walk-INF)

‘João walked to campus and Maria did the same to the café’. (do the same = to walk)

When a PP headed by *até* is combined with a manner of motion verb, it is an adjunct, as shown in (33)–(34). By contrast, when combined with *ir/ vir* (‘go/ come’), the PP behaves like a complement, as shown in (9)–(10) above.

- (33) *O que é que o João fez até à faculdade? /*
 What be-PRS.3ps that the João do-PST.3SG to to-the campus? /
Caminhou.

Walked-PST.3SG.

‘What did João do (up) to campus? / He walked’.

he PP behaves like a complement, as shown in (9)–(10) above.

- (33) *O que é que o João fez até à faculdade? /*
 What be-PRS.3ps that the João do-PST.3SG to to-the campus? /
Caminhou.

Walked-PST.3SG.

‘What did João do (up) to campus? / He walked’.

- (34) *O João caminhou até à faculdade e a Maria fez*
 The João walk-PST.3SG to to-the campus and the Maria do-PST.3SG
o mesmo até ao café. (fazer o mesmo = caminhar)
 the same to to-the café. (do the same = walk-INF)

‘João walked (up) to campus and Maria did the same (up) to the café’. (do the same = to walk)

Second, manner of motion verbs can occur without a complement, but they do not have a deictic reading when the PP is absent, in contrast to *ir/ vir* (cf. (11)–(12) and (35)).

- (35) *O João caminhou.*
 The John walk-PST3.SG
 ‘John walked’.

Third, when these verbs combine with *in x time* adverbials as exemplified by (36a) and (36b), the whole event is measured, although the combination with *para* seems less acceptable, as shown by (36b):

- (36) a. *O rapaz caminhou até à faculdade em 5 m.*
 The boy walk-PST3.SG to to-the campus in 5 min.
 ‘The boy walked to campus in 5 min’.
- b. ^{ok/?}*O rapaz caminhou para a faculdade em 5 m.*
 The boy walk-PST3.SG to the campus in 5 min.
 ‘The boy walked to campus in 5 min’.

By contrast, when a *for x time* adverbial occurs, there are some differences depending on the use of *até* or *para*. When the preposition is *até*, the temporal adverb does not measure a consequent state, diverging from what happens with *ir/ vir* verbs. To see this contrast, compare examples (37a) and (19), repeated below as (37b).

The combination of verbs like *caminhar* with such an adverbial allows two possible readings. First, we find a single-event reading in which the entity that undergoes movement reaches the destination; in this case, the two classes of verbs are similar. However, we find an additional, iterative interpretation, with *caminhar* that is not available for *ir/vir* verbs: ‘*there is a series of events of the boy walking to campus that lasts 1 hr*’..

- (37) a. *O rapaz caminhou até à faculdade durante 1h.*
 The boy walk-PST3.SG to to-the campus for 1hr.
 Lit. ‘The boy walked to campus for 1hr’.
- b. *O rapaz foi/veio até à faculdade durante 5 m.*
 The boy go/come-PST.3SG to to-the campus during 5 min.
 ‘The boy went/came to campus for 5 min’. (= he was there 5 min.)
 (= *esteve lá 5 m*)

When *for x time* adverbial occurs and the PP is headed by *para*, it only admits a single event reading, in which the entity that undergoes movement does not achieve destination, as illustrated by (38). By contrast, the predications with *ir* and *vir*, in this context, admit not only the single event reading, in which the entity that undergoes movement does not reach destination, but also a resultant state reading (cf. (20) above).

- (38) *O rapaz caminhou para a faculdade durante 1h.*
 The boy walk-PST.3SG to the campus for 1hr.
 ‘The boy walked to campus for 1hr’.

In sum, despite the syntactic differences – the PPs are adjuncts with manner of motion verbs while they are complements with *ir/ vir* – there are salient semantic similarities: a *para* headed PP allows both a reading in which the entity that undergoes movement reaches its destination and a mere directional interpretation in which the entity that undergoes movement does not achieve destination; an *até*-headed PP allows the former but not the latter.

In order to explain the different readings of predications with manner of motion verbs like *caminhar* and prepositions *para* and *até*, Leal and Oliveira (2015) compare two frameworks: a mereological approach (following Zwarts 2005) and a scalar approach (following Kennedy & Levin 2008, for degree achievements). The authors show that the former cannot deal with variable telicity of predications with *para*. By contrast, a scalar approach was found to provide a better account of the different readings of these predications.

In the next section, we argue that this proposal, with some enhancements, can explain the variable telicity of predications with verbs *ir/ vir* combined with prepositions *para/ até*. We assume that these verbs have a lexical meaning that includes the notion of path scale (Rappaport Hovav 2014) and we contend that the prepositions contribute in a different way to the specification of one of the parameters of this scale.

5. The proposal

5.1 Scale semantics as a theoretical framework

Our proposal is based on Kennedy & Levin (2008) account of degree achievements, which is an improvement over Hay, Kennedy & Levin (1999) (for a different scalar based proposal, see Civardi & Bertinetto 2015). In a nutshell, Kennedy & Levin (2008) argue that the meaning of deadjectival verbs that project degree achievements (e.g. *darken*) is a function that measures the degree of change of an entity as a result of participating in an event in relation to a scale dimension. In other words, a degree achievement verb expresses a measure of change function: (a) whose domain contains an individual and an event; and (b) whose range contains the degree that indicates a change occurred in that individual. This change is related to the property expressed by the adjectival base, as a result of the individual participating in the event described by the verb.

This measure of change function is a special case of difference function and projects an individual x and an event e in a degree that represents the amount of change verified in x in a property measured by a measure function m as a result of x participating in e . The input of this measure of change function is the minimum value of the scale that corresponds to the amount of the relevant property exhibited by x in the beginning of the event. The output is a degree that represents the positive difference between the minimum value of the scale and the value of x in the end of e . The truth conditions of a degree achievement are as follows: it applies to an object x and an event e if and only if the degree in which x changes as a result of participating in e exceeds the standard of comparison of the measure of change function. For instance, a predication with a degree achievement verb, as in *the sky darkened*, is true if and only if the entity denoted by *the sky* changes in one of its properties (the property of being dark) as a result of being involved in an event of *darkening* in such a way that, at the end of the event, this entity becomes dark.

Kennedy and Levin (2008) use this proposal to explain the different behavior of degree achievement verbs in regard to telicity. For these authors, some English verbs, like *darken*, have a default telic interpretation – the atelic interpretation is highly marked. The telic interpretation arises because these verbs are derived from closed scales adjectives, i.e., scales with maximal elements.¹² These verbs inherit the adjectival properties, namely the scale property, so they only apply if the individual x involved in the event e displays the maximal degree of the scale in the end of e . Using again the example *the sky darkened*, since *dark* is a closed scale adjective, the sentence is true if and only if, in the end of the event, the sky is *completely dark*. Along with this maximum standard interpretation, there is also the possibility of a minimum standard, or atelic interpretation. In this case, in order for the predication to be true, it is only necessary that there be some positive change in the degree displayed by x and there is no need for this degree to be the maximal one. In the case of degree achievement verbs like *darken*, the telic interpretation is preferred for pragmatic reasons. According to the principle of Interpretive Economy (Kennedy & Levin 2008), one must maximize the contribution of the conventional meanings of the linguistic elements to the computation of truth conditions of sentences. As the telic interpretation ('in the end of the event, the sky is completely dark') lexically

12. Following Kennedy & McNally (2005), one way of distinguishing open scale from closed scale adjectives is to look to the distribution of proportional modifiers, like *half*. Only the latter, but not the former, are acceptable with these modifiers. See the following examples, from Kennedy & McNally (2005: 352–353).

- (a) The glass is half full (closed scale adjective)
- (b) ??The rope is half long (open scale adjective)

entails the atelic one ('in the end of the event, the sky is darker than in the beginning of the event') the former is more informative than the latter and its cancellation depends on contextual factors (cf. Kennedy and Levin's example 'the sky darkened but it didn't become dark').

On the other hand, there are degree achievement verbs, like *widen*, which have only an atelic interpretation because they are related to adjectives that denote open-scale measure functions. These adjectives project scales without a maximal degree; therefore, the verbs that inherit this scalar structure cannot have the telic (maximal standard) interpretation – only the atelic (minimal standard) interpretation is lexically available.

Kennedy and Levin (2008), Kennedy (2012), Rappaport Hovav and Levin (2010) and Rappaport Hovav (2014), among others, argue that this scalar approach can be extended to explain the telicity of predications with other types of verbs, namely inherently directed motion verbs. In fact, there is evidence that this type of verb also has a scalar meaning. However, according to Rappaport Hovav and Levin (2010), only a few verbs lexicalize all components of the scale. Rappaport Hovav (2014) points out that, contrary to what happens with change-of-state verbs, motion verbs require an element of predication to specify one of the components of the path scale.

Kennedy and McNally (2005) claim that a scale has three components: a set of degrees, a dimension and an ordering relation. The set of degrees represents measurement values that can include (or not) minimal and maximal elements. In the case of verbs of movement, the values correspond to locations along a path. The dimension indicates the kind of measurement, for instance, height, volume, temperature, or, in the case of verbs of movement, spatial location relative to a 'reference object' (cf. Rappaport Hovav 2014). Finally, the ordering relation makes the linear order of the degrees explicit. Rappaport Hovav (2014) also argues that the ordering relation is related to some 'reference object' in the case of verbs of movement: these verbs specify if the movement is toward or away from the 'reference object'.

Following Beavers (2008), we also assume that verbs can fall under three different categories concerning durativity: they can be durative, punctual or underspecified. Punctual verbs are Minimally Complex Objects (Beavers 2008: 254), that is, they have exactly two subparts (they are related to two-point scales), whereas durative verbs are Complex Objects, that is, they have at least three subparts (they are related to at least three point scales). Underspecified verbs allow other elements of the predication (or the context) to determine whether the verb is associated with a two-point scale or with a multi-point scale.

In the next section, we show how the theoretical framework just described can be used to account for the patterns exhibited by *ir/ vir* in combination with *para/ até*.

5.2 Analysis of the data

In order to explain the data in EP using the model explained above, we assume that inherently directed motion verbs *ir/ vir* have an underlying scalar structure just like degree achievements. That structure is (under)specified as follows: (a) the measuring dimension is specified (the path's dimension); (b) the set of degrees (i.e. whether there is a maximum value in the scale) is not fully specified; (c) the ordering relation (i.e. distance from a certain point) is not fully specified either. In the last two cases (set of degrees and ordering relation), it is necessary for an element of the predication to denote the reference object in order to define the relevant path (cf. Rappaport Hovav 2014).

We argue that PPs with *para* and *até* contribute in a distinct way to the specification of one of the parameters of path scales (see, although with distinct motivations, Fleischhauer & Gameschlag 2014). PPs headed by the preposition *para* only specify the ordering relation; that is, *para* operates on the ordering relation by defining an arbitrary degree (an arbitrary location) in the path scale. In this case, the interpretation of the predication includes the existence of an individual *x* participating in an event *e* that is approaching the arbitrary degree defined by PP-*para* in the course of *e*. By contrast, the preposition *até* operates on the set-of-degrees parameter and denotes a maximal element that is contextually relevant. As a consequence, the PP-*até* transforms the scale projected by the motion verb to a closed scale.

From an aspectual point of view, PPs with *para* have the atelic interpretation as their default interpretation, since *para* does not define a maximal degree (only an arbitrary degree) and, therefore, the scale projected by the verb is an open scale. This means that there is only a change of location (of degree) of individual *x* in the course of the event, but at the end of the event, *x* cannot achieve the maximal degree, as this degree does not exist. Thus, in (39a) and (39b), the default interpretation of PP-*para* is merely directional, that is, non-culminative.

- (39) a. *O rapaz foi para a faculdade.*
 The boy go-PST.3SG toward the campus.
 'The boy went toward campus'.
 b. *O rapaz veio para a faculdade.*
 The boy come-PST.3SG toward the campus.
 'The boy came toward campus'.

However, in a context in which telicity is forced, as with an *in x time* adverbial, an aspectual shift is necessary (which explains why some informants consider (40) slightly odd), as there is incompatibility between the information provided by the PP – a mere (open) scale orientation – and the adverbial, which requires a telic

predication, that is, a closed scale. In this case, the arbitrary location (arbitrary degree on the scale) denoted by PP-*para* must be reinterpreted as a maximal degree, similar to PP-*até*.

- (40) *ok/?* *O rapaz foi/veio para a faculdade em 5 m.*
 The boy go/come-PST.3SG to the campus in 5 min.
 ‘*ok/?*The boy went/came to campus in 5 min.’

As far as *até* is concerned, its interpretation is the telic one, since *até* defines a maximal degree and ensures that the scale associated with the verb is contextually closed. This means that there is an individual *x* participating in an event *e* and, at the end of *e*, *x* is located in the maximal degree of the scale – the maximal location (cf. (41)).

- (41) *O rapaz foi/veio até à faculdade (em 5 m).*
 The boy go/come-PST.3SG to-to-the campus (in 5 min).
 ‘The boy went/came to campus (in 5 min.)’

In all cases above, verbs *ir/ vir* are interpreted as durative verbs, which means that they are associated with multi-point scales. However, when these verbs combine with PP-*até* and a *for x time* adverbial, the situation changes. In these cases, with PP-*até*, there is only a reading of measurement of the consequent state (cf. (42)), whereas with PP-*para*, there is some ambiguity between a reading of measurement of the preparatory process of the event (first reading of (43): *estive a ir/ vir para a faculdade durante 5 m*) and a reading of measurement of the consequent state, similar to what happens with PP-*até* (second reading of (43): *estive na faculdade 5 m*).

- (42) a. *O rapaz foi até à faculdade*
 The boy go-PST.3SG to-to-the campus
durante 5 m. (= esteve na faculdade 5 m)
 during 5 min.
 ‘The boy went to campus for 5 min. (= he stayed on campus for 5 min.)’
 b. *O rapaz veio até à faculdade*
 The boy come-PST.3SG to-to-the campus
durante 5 m. (= esteve na fac. 5 m)
 during 5 min.
 ‘The boy came to campus for 5 min. (= he stayed on campus 5 min.)’
 (43) a. *O rapaz foi para a faculdade durante*
 The boy go-PST.3SG to the campus during
5 m. (= esteve a ir para a faculdade durante 5 m./estive na faculdade 5 m.)
 5 min.
 ‘The boy went toward/to campus for 5 min. (= he was going toward campus for 5 m./ he stayed on campus for 5 min.)’

- b. *O rapaz veio para a faculdade durante*
 The boy come-PST.3SG to the campus during
 5 m. (= *esteve a vir para a faculdade durante 5 m./esteve na faculdade 5 m.*)
 5 min.
 ‘The boy came toward/to campus for 5 min. (= he was coming toward
 campus for 5 min./ he stayed on campus for 5 min.)’

In the examples in (42a-b), predications with verbs *ir/ vir* are interpreted as non-durative, that is, as evoking two-point scales. As these predications are interpreted as non-durative (*ir/ vir* behave like achievement verbs), there is no process phase of the event to be measured by the temporal adverbial. The PPs denote the maximal degree and verbs are interpreted as transitions between a state of not being at the destination and the state of being at the destination. Therefore, the adverbial can only measure the resultant state of the event, that is, the time interval during which the individual *x* participating in the event *e* remains in destination after the end of *e*. This reading is possible not only with PP-*até*, but also with PP-*para*, provided that the predications are interpreted as non-durative and the arbitrary degree defined by PP-*para* is reinterpreted as a maximal degree (cf. second reading of (43a) and (43b)). In other words, this resultant state reading of predications with PP-*para* is the product of an aspectual shift.

6. Concluding remarks

The main objective of this chapter was to put forward a proposal that accounts for the semantics of inherently directed motion verbs (Levin 1993) *ir* (‘go’) and *vir* (‘come’) combined with PPs with the thematic role of Goal, headed by the prepositions *para* (‘to/ toward’) and *até* (‘to’) in EP. One of the main issues discussed was the possibility of telic and/ or atelic interpretations of these predications.

To do so, we built on the notion of path scale (e.g. Hay, Kennedy & Levin 1999; Kennedy & McNally 2005; Kennedy & Levin 2008) and we assumed that, like degree achievements, motion verbs have an underlying scalar structure, which is underspecified (Rappaport Hovav & Levin 2010) in the following way: the measuring dimension is specified (paths) but the set of degrees, namely whether there is a maximum value in the scale, and the ordering relation (proximity, or not, to a certain point) are not fully specified. Furthermore, following Leal & Oliveira’s (2015) hypothesis concerning the combination of *para* and *até* with manner of motion verbs, we argued that these PPs contribute in a distinct way to the specification of one of the parameters of path scales (based on Fleischhauer & Gameschlag 2014): *para* specifies the ordering relation (approach to a certain point defined by the

PP-*para*), whereas *até* operates on the parameter of set of degrees, and denotes a maximum element that is contextually relevant, and, thus, transforms the scale projected by the verb into a closed scale.

So, this proposal assumes that inherently directed motion verbs *ir/ vir* have a lexical meaning that can be represented as a partially specified scale (cf. Rappaport Hovav & Levin 2010) and that the PPs headed by *para* and *até*, which are complements of these verbs, contribute in different ways to specify some parameters of the scale: (i) *para* determines the ordering relation in the scale projected by the verb and (ii) *até* determines the set of degrees and denotes the maximal element of the scale projected by the verb.

We have also showed that these predications are generally durative (cf. examples (13) and (14), with *in x time* adverbial), but under some circumstances they can have a non-durative reading (cf. examples (19) and (20), with *for x time* adverbial).

To conclude, we argue that the contribution of PPs headed by *para* and *até* in EP to the definition of the aspectual profile of the predications that denote events of movement can be accounted by scalar semantics. This explanation can accommodate the data concerning not only verbs of manner of motion, but also inherently directed motion verbs *ir* and *vir*, and relates these verbs of movement to other types, namely degree achievement verbs.

References

- Beavers, John. 2008. Scalar complexity and the structure of events. In *Event Structures in Linguistic Form and Interpretation*, Johannes Dölling, Tatjana Heyde-Zybatow & Martin Schäfer (eds), 245–265. Berlin: Mouton de Gruyter. doi:10.1515/9783110925449.245
- Civardi, Eugenio & Bertinetto, Pier Marco. 2015. The semantics of degree verbs and the telicity issue. *Borealis: An International Journal of Hispanic Linguistics* 4(1): 57–77. doi:10.7557/1.4.1.3398
- Cordeiro, João & Brazdil, Pavel. 2017. Estudo empírico da linguagem humana com recurso à Web. Paper presented at the workshop *Verbos e Preposições*, Porto, May.
- Dowty, David R. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel. doi:10.1007/978-94-009-9473-7
- Fleischhauer, Jens & Gameschlag, Thomas. 2014. We're going through changes: How change of state verbs and arguments combine in scale composition. *Lingua* 141: 30–47. doi:10.1016/j.lingua.2013.01.006
- Hay, Jennifer, Kennedy, Christopher & Levin, Beth. 1999. Scalar structure underlies telicity in 'degree achievements'. In *Proceedings of SALT IX*, Tanya Matthews & Devon Strolovitch (eds), 127–144. Ithaca NY: CLC Pub.
- Kennedy, Christopher. 2012. The composition of incremental change. In *Telicity, Change, State: A Cross-categorical View of Event Structure*, Violeta Demonte & Louise McNally (eds), 103–121. Oxford: OUP. doi:10.1093/acprof:oso/9780199693498.003.0004

- Kennedy, Christopher & Levin, Beth. 2008. Measure of change: The adjectival core of degree achievements. In *Adjectives and Adverbs: Syntax, Semantics and Discourse*, Louise McNally & Christopher Kennedy (eds), 156–182. Oxford: OUP.
- Kennedy, Christopher & McNally, Louise. 2005. Scale structure and the semantic typology of gradable predicates. *Language* 81: 345–381. doi:10.1353/lan.2005.0071
- Leal, António & Oliveira, Fátima. 2008. Subtipos de verbos de movimento e classes aspectuais. In *Textos Seleccionados do XXIII Encontro da Associação Portuguesa de Linguística*, Maria Lobo & Maria Antónia Coutinho (eds), 287–298. Lisboa: Associação Portuguesa de Linguística.
- Leal, António. & Fátima Oliveira. 2015. Verbos de movimento, preposições direcionais e escalas. *Revista da Associação Portuguesa de Linguística* 1: 353–366.
- Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago IL: The University Chicago Press.
- Moens, Marc. 1987. *Tense, Aspect and Temporal Reference*. PhD dissertation, University of Edinburgh.
- Rappaport Hovav, Malka. 2014. Building scalar changes. In *The Syntax of Roots and the Roots of Syntax*, Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds), 259–281. Oxford: OUP. doi:10.1093/acprof:oso/9780199665266.003.0012
- Rappaport Hovav, Malka & Levin, Beth. 2010. Reflections on manner/result complementarity. In *Syntax, Lexical Semantics, and Event Structure*, Malka Rappaport Hovav, Edit Doron & Ivy Sichel (eds), 21–38. Oxford: OUP. doi:10.1093/acprof:oso/9780199544325.003.0002
- Vendler, Zeno. 1957. Verbs and times. *Philosophical Review* 66:143–160. doi:10.2307/2182371
- Zwarts, Joost. 2005. Prepositional aspect and the algebra of paths. *Linguistic and Philosophy* 28(6): 734–779. doi:10.1007/s10988-005-2466-y

Name index

A

Abouda, Lotfi 65, 79, 328
Abraham, Werner 3, 4, 5, 16,
18, 191, 208, 210, 211
Aijmer, Karin 119, 144, 148, 161,
166–169, 176, 182
Aikhenvald, Alexandra Y.
3, 4, 5, 16, 85, 104, 118, 142,
149, 161, 166, 182, 186, 187, 208,
236, 240, 252, 258
Akatsuka, Noriko 227, 234
Alcázar, Asier 221, 229–230,
234
Ambrasas, Vytautas 241, 258
Ameka, Felix 217, 219, 234
Amon, Marri 8, 41, 62
Anand, Pranav 199, 201, 208
AnderBois, Scott 197, 208
Anderson, John M. 256, 258
Anderson, Lloyd B. 3, 16, 85,
88–89, 93, 104, 118–119, 142,
149, 161, 256
Anderson, Stephen R. 190, 208
Anscombre, Jean-Claude
316, 328
Antoniou, Maria 313, 328
Arkadiev, Peter 241, 258
Asher, Nicholas 294, 304
Austin, John L. 167, 197, 208
Avilova, Natalia 289, 304
Ayoun, Dalila 1, 311, 327–328,
329
Azzopardi, Sophie VII, 3, 16,
20, 28, 38, 66–67, 79, 81, 213,
235

B

Babko-Malaya, Olga 289, 304
Barbet, Cécile VII, 43, 45–49,
51, 54–55, 57–58, 61, 64, 188,
208
Barceló, Gerard Joan 69, 79

Bat-Zeev Shyldkrot, Hava
317, 324, 328
Baude, Olivier 329
Bayer, Josef 201, 208, 209, 211
Beavers, John 350, 354
Becher, Gérard 17
Bende-Farkas, Agnès 299–300,
304
Benveniste, Émile 66, 79, 283,
305
Berdasco-Gancedo, Yolanda
7, 10, 109, 114, 142
Berthonneau, Anne-Marie
310, 313, 316, 328
Bertinetto, Pier Marco 5, 16,
281, 285, 305, 307, 348, 354
Beysade, Claire 228, 235
Biber, Douglas 113, 142, 330
Bilger, Mireille 310, 328
Binnick, Robert, I. 5, 16, 279,
311, 328
Blumenthal, Peter 316, 328
Blyth, Carl 327, 329
Boas, Hans C. 185, 309, 329,
330
Bondarko, Alexander 290, 305
Boogaart, Ronny 312, 329
Borer, Hagit 281, 286, 290,
305, 355
Borik, Olga 289, 305
Borkin, Ann 243, 256, 258
Bourova, Viara 213, 235
Bouveret, Myriam 309, 329, 333
Boye, Kasper 4, 5, 16, 83, 88, 104,
114–115, 122, 142, 239–240,
244, 258
Brazdil, Pavel 337, 354
Bres, Jacques VII, 3, 5, 8–9,
16, 20, 38, 65, 67–69, 78, 79,
213, 235
Brinton, Laurel J. VII, 167, 182
Brisard, Frank 83–84, 104, 162

Broccias, Cristiano 314, 329
Brugman, Claudia 186–188,
190, 196, 208
Brysaert, Marc 331
Bybee, Joan 2–3, 16, 23, 38, 46,
62, 70, 80, 143, 147–148, 150,
161, 262–264, 279, 282–283,
305, 310, 315, 329
Bylsma, Lauren 238
Bürgel, Christoph 317–319, 332

C

Cabré, M. 113, 142
Camus, Albert 14, 59, 62, 281,
305, 307
Caponigro, Ivano 222, 224, 235
Cappeau, Paul 310, 328
Cardinaletti, Anna 193, 208
Carlson, Lauri 312, 329
Carretero, Marta 7, 10, 18,
105–106, 109, 122, 142–144,
259
Carroll, James M. 215, 235
Cassel, Daniel 2, 18
Cassel, Jean-Christophe 2, 18
Castelfranchi, Cristiano 230,
236
Caudal, Patrick 283, 305, 331
Celle, Agnès 1, 2, 4–5, 7, 12, 16,
17, 19, 38, 42, 53, 62, 86, 103,
104, 110, 118, 143, 213–218,
220–222, 227, 229–231, 233,
235, 237–238
Chafe, Wallace 3, 16, 85, 104–
105, 142, 146, 161, 166, 182, 187,
208, 210
Chilton, Paul 83–84, 91–95,
101–104, 105
Chocheyras, Jacques 317, 329
Chojnicka, Joanna 120–122, 143
Chung, Kyung-Sook 196, 208
Civardi, Eugenio 348, 354

- Clark-Polner, Elizabeth 237
 Clore, Gerald L. 237
 Coates, Jennifer 87, 105, 126, 143
 Coleman, Timothy 240, 259
 Comrie, Bernard 2, 5, 17, 66, 80, 283, 294, 305, 311, 329
 Confais, Jean-Paul 19, 38
 Coniglio, Marco 12, 186, 192, 201, 206, 208
 Cordeiro, João 337, 354
 Cornillie, Bert 86, 105, 106–107, 144, 147–149, 160, 161, 162, 188, 209
 Croft, William 7, 17, 282, 288, 305, 314, 329
 Cruschina, Silvio 86, 105
 Cruse, Alan D. 314, 329
 Csirmaz, Aniko 283, 296, 305
 Culioli, Antoine 165, 174–175, 181, 182–183
 Cuzzolin, Pierluigi 242, 259
- D**
- Dahl, Östen 2, 6, 16–17, 41, 57, 63, 264, 269–270, 279, 282, 305, 307
 Damourette, Jacques 19, 38, 65, 68–69, 80
 Darwin, Charles 217, 235
 Davis, Christopher 197, 209
 Davis, Henry 211
 Diewald, Gabriele 3–4, 17, 85, 105, 192, 209, 260, 265, 279
 Debaisieux, Jeanne-Marie 310, 329
 de Haan, Ferdinand 3, 5, 17, 119, 143, 147–148, 160, 161, 187, 209
 De Knop, Sabine 309, 329
 DeLancey, Scott 4, 17, 220, 230, 235
 Delfitto, Denis 5, 16
 de Mulder, Walter 68, 70, 80, 147–148, 161, 312, 331, 333
 Dendale, Patrick 44, 63, 86, 88, 105, 106, 147–149, 153–154, 161, 162, 163, 187–188, 209, 213, 216, 235, 236
 den Dikken, Marcel 220, 235
 Depraetere, Ilse 281, 305
 de Saussure, Louis 84, 91–92, 103, 104, 105, 106, 188, 237, 311, 332
 Desclés, Jean-Pierre 182, 313, 329
 Desmets, Marianne 227–228, 236
 de Swart, Henriëtte viii, 6, 17, 282–283, 285, 298, 305, 307, 312, 332
 Detges, Ulrich 70, 80, 313, 315–316, 322, 329
 Deulofeu, Henri-José 310, 329
 De Vlam, Cornelis 238
 De Wit, Astrid 7, 17
 Dickey, Stephen M. 291, 305
 Diessel, Holger 258, 315, 329
 Dik, Simon C. 43, 252, 282, 305
 Diller, Anne-Marie 213, 216, 236
 Dion, Nathalie 65, 81
 Di Sciullo, Anna-Maria 289, 305
 Di Vito, Nadine O'Connor 313, 329
 Diwersy, Sascha 9, 65, 72, 80, 317, 332
 Do-Hurinville, Danh-Thành 313, 329
 Dowty, David R. 340–341, 354
 Du Bois, John 4, 17, 231
 Duffley, Patrick J. 243, 256, 259
 Dugua, Céline 229
 Détrie, Catherine 331
- E**
- Eckardt, Regine 194, 209
 Egg, Markus 193, 196, 201, 209
 Ellis, Nick 315, 329
 Elman, Jeffrey L. 330
 Enjalbert, Patrice 17
 Erelt, Mati 46, 63
 Eshkol-Taravella, Iris 310, 329
 Esteve Ramírez, Francisco 113, 143
 Evert, Stefan 72, 80, 183
- F**
- Facques, Bénédicte 313, 330
 Falaise, Achille 80, 318, 330
 Faller, Martina 4, 17, 120, 143, 186, 196–199, 203, 209
 Farkas, Donka 298, 305
 Fauconnier, Gilles 94, 105
 Felser, Claudia 243, 259
 Feng, Hanbing 229, 236
 Fernández del Moral, Javier 113, 143
 Fillmore, Charles 218, 221, 236, 306, 314, 330
 Fleischhauer, Jens 351, 353, 354
 Fleischman, Suzanne 1–2, 17, 19, 38, 65–66, 68, 78, 80, 151, 320, 338
 Fletcher, William H. 169, 183
 Floyd, Rick 147–148, 161, 237
 Floyd, Simeon 213, 217, 233, 237
 Forker, Diana 269, 273, 275–276, 279
 Forsyth, John 294, 305
 Fox, Renata 113, 143
 Frajzyngier, Zygmunt 143, 147–148, 161
 Fritz, Thomas A. 19, 22, 38
 Furkó, Péter 282, 305
- G**
- Gameschlag, Thomas 351, 353, 354
 Garachana Camarero, Mar 161
 Gardies, Jean-Louis 67
 Garey, Howard B. 311, 330
 Garrett, Edward 190, 196, 209
 Gautier, Antoine 227–228, 236
 Gavarró, Anna 148, 161
 Gehrke, Berit vii, 289, 305
 Giannakidou, Anastasia 220, 235
 Gide, André 51, 63
 Gillies, Anthony S. 4, 18, 188, 190, 195, 197, 209
 Gimpel, Jean 44, 63
 Ginzburg, Jonathan 222, 236
 Givón, Talmy 83, 84, 88, 89, 103–104, 105, 111, 143, 187, 209, 330
 Glovinskaja, Marina 294, 305
 Goldberg, Adele 7, 17, 236, 314, 326, 330
 González, Montserrat vii, 148, 161

- Gosse, André 23, 38
 Gosselin, Laurent VII, 7, 17, 283, 305
 Gougenheim, Georges 69, 80
 Greco, Paolo 242, 257, 259
 Grevisse, Maurice 23, 38
 Grice, Paul 191, 197
 Gries, Stefan T. 73–74, 80, 310, 314–315, 319, 327, 330, 332
 Gronemeyer, Claire 240, 259
 Gross, Maurice 183, 316, 330
 Grønn, Atle 284, 294–295, 305
 Guentchéva, Zlatka 3, 5–6, 17, 161, 228, 235, 236, 305, 313, 329
 Guillemin-Flescher, Jacqueline 282, 306
 Gunlogson, Christine A. 223, 236
 Gutzmann, Daniel 193, 209
- H**
 Habicht, Külli 43, 63
 Hacquard, Valentine 199, 208
 Haegeman, Liliane 186, 206, 208, 209, 210, 211
 Haillet, Pierre-Patrick 213, 236
 Hamblin, Charles L. 202, 210
 Harder, Peter 4, 16, 88, 104, 115, 142, 240, 258
 Hardie, Andrew 72, 80
 Hare, Mary 317, 330
 Haumann, Dagmar 193, 210
 Hay, Jennifer 348, 353, 354
 Heim, Johannes 206, 210
 Heine, Bernd 32, 38, 62, 68, 80, 183
 Helbig, Gerhard 192, 210
 Hennemann, Anja 118, 143
 Henry, Albert 70, 80
 Hentschel, Elke 192, 210
 Herzog, Christian 313, 330
 Heycock, Caroline 203, 210
 Hilpert, Martin 73–74, 80
 Hoeksema, Jack 221, 236
 Hoey, Michael 172, 183
 Hoffmann, Lothar 143
 Hoffmann, Sebastian 165, 183
 Hofmann, Johann 242, 249, 259
 Holmes, Janet 168, 176, 183
 Holvoet, Axel VII, 240, 259
- Hopper, Paul J. 68, 80, 283, 306, 312, 330
 Hriba, Linda 329
 Huddleston, Rodney 2, 17, 126, 143, 216, 218, 220–221, 224, 228, 236
 Hunston, Susan 326, 330
- I**
 Imbs, Paul 311, 330
 Izvorski, Roumyana 186, 196, 210
- J**
 Jacobs, Joachim 201, 210
 Jacobsen, William H. 185, 210
 Janda, Laura 290, 306
 Jasionytė, Erika 241, 259
 Jaszczolt, Kasia M. 84, 91, 104, 105, 106
 Jaxontov, Sergei 263, 279
 Jespersen, Otto 2, 17
 Johanson, Lars 5, 17
 Jugnet, Anne 234, 235
- K**
 Kaltenböck, Gunther VIII, 165, 168–169, 174, 176–178, 180, 183
 Kamp, Hans 312, 330
 Kaplan, David 201, 210, 234
 Kardos, Éva A. 299, 301, 306
 Karttunen, Lauri 202, 210
 Kaufman, Daniel 190, 210
 Kay, Paul 218, 236, 314, 326, 330
 Kennedy, Christopher 335, 348–350, 353–355
 Keupdjio, Hermann 210
 Kibrik, Andrej 269, 279
 Kiefer, Ferenc 283, 296, 298, 303, 306
 Kilgi, Annika 43, 63
 Kiss, Katalina É. 283, 296, 298–299, 301, 303, 305–306
 Kleiber, Georges 38, 310, 313, 316, 328
 Klein, Wolfgang 5, 17, 281, 283, 306
 Klum, Arne 311, 330
 Koops, Christian 171, 183
 Korotkova, Natalia 5, 17, 186, 197, 199, 210, 221, 236
- Kratzer, Angelika 196, 201, 210
 Krifka, Manfred 286, 306, 312, 330
 Kripke, Saul A. 67, 80
 Kroeger, Paul 190, 218
 Kronning, Hans 41, 43–46, 48–51, 54–58, 60–61, 63, 86, 105, 148, 162
 Kross, Jaan 52–55, 58, 63
 Kuhn, Julia 329
 Kulikov, Leonid 5, 16
 Kuteva, Tania 6, 68, 70, 80
 Kuteva, Tania 46, 63, 68, 70, 80
 Kärkkäinen, Elise 169, 183
- L**
 Labeau, Emmanuelle 68–69, 78, 79, 313, 330, 331
 Labelle, Marie 313, 317, 331
 Laca, Brenda VII, 86, 105, 148, 161, 307
 Lakoff, Robin 167, 183
 Lambrecht, Knud 218, 236
 Lampert, Günther 120, 143
 Langacker, Ronald W. 83–84, 86–87, 91, 93–94, 100, 103, 105, 145–146, 153, 162
 Lansari, Laure 1–2, 17, 219, 227, 234, 235, 237
 LaPolla, Randy J. 6, 18, 85, 107
 Larreya, Paul VII, 126, 143
 Larrivée, Pierre 313, 330–331
 Lascarides, Alex 294, 304
 Lavid, Julia 116, 138, 143
 Lavine, E. 240, 259
 Lay, Marie-Hélène 80
 Leal, António 8, 15, 335–336, 346, 348, 353, 355
 Lee, David 183
 Legallois, Dominique 309, 329, 331, 333
 Leinonen, Marja 294, 306
 Leiss, Elisabeth 3, 5, 16, 18, 191, 208, 210–211
 Lemmens, Maarten 4, 18
 Le Querler, Nicole 148, 162
 Letnes, Ole 193, 210
 Leumann, von Manu 242, 249, 259
 Levin, Beth 335–336, 348–350, 353–355

- Levshina, Natalia 318, 331
 Li, Charles N. 122, 143
 Lichtenberk, Frantisek 116, 143
 Lindschouw, Jan 65, 80, 105
 Lindstedt, Jouko 262, 266, 279
 Littell, Patrick 213–214, 216–217, 231, 233, 236
 Lière, Audrey 65, 81
 Lohmann, Arne 171, 183
 Lopez, Cédric 331
 Lowe, John M. 257, 259
 Luscher, Jean-Marc 106, 311, 331
 Lyons, John 2, 18, 126, 143, 197, 210, 255, 259
 Lévi-Strauss, Claude 55, 63
 L'Hôte, Emilie 234–235
- M**
- Macaulay, Monica 186–188, 190, 196, 208
 Mackie, Scott 186, 212
 Maisak, Timur VIII, 269, 279
 Mallaeva, Zulajxat 264, 269, 279
 Manning, Liliann 1, 18
 Marandin, Jean-Marie 228, 235
 Martin, Robert 67, 81
 Martin du Gard, Roger 44, 63
 Martinez, Ron 318, 331
 Martínez-Egido, José J. 109, 123, 144
 Marín Arrese, Juana I. VII, 7, 9, 83, 84, 86, 91, 99, 103, 105–107, 120, 143, 144, 259
 Maslov, Jurij 290, 306
 Mathet, Yann 17, 215
 Matthewson, Lisa 186–189, 196–198, 203, 211, 212, 236
 Maurel, Denis 329
 Mañoso Pacheco, Lidia 110, 137, 144
 McCawley, James 283, 306
 McCready, Eric VIII, 186, 188, 196, 211
 McNally, Louise 349–350, 353–355
 McRae, Kevin 330
 Metslang, Helle VIII, 43, 46, 53, 63
 Meyer, Wulf Uwe 223, 230, 232, 236
- Miceli, Maria 230, 236
 Michaelis, Laura VIII, 2, 7, 18, 218, 229, 236, 282, 285–286, 306, 315, 326, 330, 331
 Mihas, Elena 230, 236
 Mithun, Marianne 3, 18
 Modicom, Pierre-Yves 194, 211
 Moens, Marc 282, 306, 341, 355
 Molendijk, Arie 283, 307, 312, 328, 331
 Mollica, Fabio 329
 Morency, Patrick 84, 106
 Mortelmans, Tanja VIII, 3–4, 18, 153, 162, 331
 Mortillaro, Marcello 237
 Moïse, Claudine 331
 Murray, Sarah 197, 211, 221, 229–230, 237
 Mursell, Johannes 7, 11, 185, 193, 196, 201, 209
 Mushin, Ilana 4, 18
- N**
- Napoli, Donna Jo 221, 236
 Nedjalkov, Vladimir 16, 263–264, 279
 New, Boris 318, 331
 Newman, John 321, 332
 Nichols, Johanna 3, 16, 104–105, 142, 161, 166, 182, 187, 208, 210
 Nikolaeva, Irina 256, 259
 Nilsson-Ehle, Hans 45, 63, 67–68, 81
 Noonan, Michael 252, 259
 Norcliffe, Elisabeth 237
 Noël, Dirk VIII, 240, 259, 322
 Nuyts, Jan 4, 18, 114, 144, 147–148, 155, 162, 163, 187, 211, 243, 259
- O**
- Obenauer, Hans-Georg 201, 208
 Ogata, Norry 186, 196, 211
 Oliveira, Fátima 15, 335–336, 346, 348, 355
 Oliveira, Teresa 86, 106
 Ortony, Andrew 223, 234, 237
 Osa-Gómez, Adriana 210
 Otanes, Fe 186, 189–191, 211
- Ø**
- Ørsnes, Bjarne 240, 259
- P**
- Padučeva, Elena V. 284, 294, 306
 Pagliuca, William 2, 16, 23, 38, 62, 80, 147, 161, 279, 283, 305
 Palander-Collin, M. 166, 183
 Pallier, Christoph 331
 Palmer, Frank R. 2–4, 18, 86–87, 106, 146, 155, 162, 188, 211
 Panckhurst, Rachel 318, 331
 Patard, Adeline 309, 313, 331
 Peirce, Charles S. 88, 106
 Pelyvás, Péter 145–146, 153, 162
 Perkins, Revere 2, 16, 23, 38, 62, 80, 147, 161, 279, 283, 305
 Pesetsky, David 221, 237
 Peterson, Tyler 186, 197, 211, 212, 218, 220, 228, 234, 235, 236, 237
 Pichon, Édouard 19, 38, 65, 68–69, 80
 Picoche, Jacqueline 43, 49, 64
 Pinker, Steven 217, 237
 Piñon, Christopher 299–300, 306
 Plungian, Vladimir 4, 18, 42, 44, 61, 64, 83, 85, 106, 107, 146, 163, 261, 263, 279, 282, 290, 306
 Poplack, Shana 65, 81
 Potts, Christopher 193, 197, 209, 211, 221, 233, 237
 Poutsma, Hendrik 2, 18
 Prillop, Külli 43, 63
 Prior, Arthur 2, 18
 Pullum, Geoffrey 2, 17, 216, 218, 220–221, 224, 228, 236
- Q**
- Quaglio, Paulo 318, 332
- R**
- Ramchand, Gillian 289, 307
 Ranger, Graham VIII, 7, 11, 165, 177, 183
 Rappaport Hovav, Malka 348, 350–351, 353–355

- Razanovaitė, Auksė 241, 259
 Rebotier, Aude 313–314, 332
 Reese, Brian 228, 237
 Reichenbach, Hans 45, 64
 Reisenzein, Rainer 215, 236, 237
 Remberger, Eva-Maria 86, 105
 Repp, Sophie 194, 196, 211
 Rescher, Nicholas 67, 81
 Rett, Jessica 221, 229–230, 237
 Revaz, Françoise 19, 38
 Ricci, Claudia 20, 38
 Rice, Sally 321, 332
 Ritz, Marie-Eve 262, 279
 Rizzi, Luigi 201, 211
 Rocci, Andrea 213, 237
 Roche, Mathieu 331
 Rohde, Hannah 222–223, 225, 229, 232, 237
 Rohrer, Christian 312, 330
 Rooryck, Johan 187–188, 211
 Ross, John R. 167, 183
 Rossari, Corinne 8, 19–20, 22, 25, 37–38
 Rullmann, Hotze 186, 211–212
 Ruskan, Anna 118, 144
 Römer, Ute 310, 314, 329, 332
- S
- Sahoo, Kalyanamalini 4, 18
 Salimov, Khangerej 269, 279
 Sanders, José 90, 106
 San Roque, Lila 213, 217, 233, 237
 Sarrazin, Sophie 3, 16, 66–67, 79, 81
 Sasse, Hans-Jürgen 311, 332
 Schachter, Paul 186, 189–191, 211
 Schenner, Mathias 203, 211
 Scherer, Klaus 215, 237
 Scherf, Nathalie 192, 211
 Schlenker, Philippe 201, 212
 Schmid, Beatrice 148, 162
 Schmitt, Norbert 318, 331
 Schützwohl, Achim 236
 Searle, John R. 125, 144, 193, 197, 212, 238
 Sentí, Andreu 8, 10, 145–148, 150–153, 155–156, 159–160, 162–163
 Siemund, Peter 217, 237
- Siepmann, Dirk 310–311, 317–319, 332
 Siminiciuc, Elena 20, 38
 Simon-Vandenberg, Anne-Marie 119, 144, 168–169, 177, 183
 Simons, Mandy 189, 212
 Skrovec, Marie 65, 79
 Slabakova, Roumyana 289, 305
 Smirnova, Alla V. 122, 144
 Smirnova, Anastasia 86, 106
 Smirnova, Elena 3–4, 17, 85, 105, 122, 260, 279
 Smith, Carlota 5, 14, 18, 281–283, 307, 312, 332
 Smith, Nicholas 183, 305
 Sode, Frank 203, 211
 Souvay, Gilles 80
 Speas, Margaret 190, 197, 209, 212
 Spooen, Wilbert 90, 106
 Spraunienė, Birutė 241, 259
 Sprouse, Jon 222, 224, 235
 Squartini, Mario 83–90, 106, 146–149, 151, 157, 160, 163, 285, 307
 Stalnaker, Robert 191, 193, 212
 Starke, Michal 193, 208
 Stathi, Katerina 119, 122, 144, 240, 260
 Steedman, Mark 282, 306
 Stefanowitsch, Anatol 310, 315, 327, 330, 332
 Stein, Achim 318, 332
 Stein, Nancy L. 218, 237
 Sten, Holger 311, 332
 Sthioul, Bertrand 91–92, 103, 106
 Stivers, Tanya 217, 237
 Stubbs, Michael 313, 332
 Sulejmanov, Jakub 269, 276, 279
 Sundell, Lars-Göran 19, 38
 Svenonius, Peter 289, 307
 Sweetser, Eve 87, 106
 Szantyr, Anton 242, 249, 259
- Š
- Šinkūnienė, Jolanta 118, 144, 241, 259
 Šmelev, Aleksej 290, 307
- T
- Talmy, Len 87, 106
 Talmy, Leonard 146, 287, 307
 Tammsaare, Anton Hansen 56, 64
 Tan, Jennifer 7, 11, 185, 189–190, 197, 212
 Tasmowski, Liliane 88, 106, 148–149, 161, 163, 187, 209, 236, 243, 259
 Tatevosov, Sergei 264–266, 279
 Taylor, John R. 322, 332
 Tellier, Isabelle 329
 Tenny, Carol 190, 212, 286, 307
 Thompson, Sandra A. 167, 183
 Thurmair, Maria 191–192, 212
 Tognini-Bonelli, Elena 313, 332
 Tonhauser, Judith 189, 212
 Torró, Jaume 150, 163
 Tournadre, Nicolas 85, 107, 282, 307
 Treikelder, Anu 7–8, 41–42, 64
 Tsangalidis, Anastasios VIII, 4, 17
 Tsertsvadze, Ilja 269, 279
 Turner, Mark 94, 105
 Turpin, Danielle 65, 81
 Tutin, Agnès 233, 237
- U
- Ultan, Russell 2, 18
 Urmson, James Opie 166–167, 183
 Urquhart, Alasdair 67, 81
 Usonienė, Aurelia 7–8, 12, 118, 143, 144, 239, 241, 259
- V
- Van Bogaert, Julie 166, 183
 van den Berg, Helma 261, 280
 van der Auwera, Johan VII, 4, 16, 18, 42, 44, 61, 64, 83, 105–107, 143–144, 146, 162–163
 van der Klook, Jozina 195, 210
 Vanderveken, Daniel 193–194, 212
 Van Dijk, Wilco W. 234, 237
 van Schooneveld, Cornelis Hendrik 290, 307
 van Wassenhove, Virginie 2, 18
 Vater, Heinz 19, 38

- Vatrican, Axelle 84, 86, 107,
122, 144
- Vendler, Zeno 6, 18, 340, 355
- Verhees, Samira 7–8, 13, 261,
265, 280
- Verine, Bertrand 331
- Verkuyl, Henk J. 286, 305, 307
- Veronis, Jean 331
- Verspoor, Marjolijn 243,
259–260
- Vet, Co 284, 307, 313, 328, 333
- Vetters, Carl VIII, 42–43,
45–49, 51, 54–55, 57–58, 61, 64,
65, 81, 148, 163, 311, 331, 333
- Vincent, Nigel 7–8, 12, 239
- Vingerhoets, Ad 217, 238
- von Fintel, Kai 4, 18, 188, 190,
195, 197, 209
- W
- Wai-Man Lam, Zoe 210
- Waldie, Ryan, Peterson 186,
197, 212
- Waugh, Linda R. 313, 333
- Weinrich, Harald 312, 333
- White, Peter R. 103, 107, 110,
123, 144
- Wiemer, Björn VIII, 86, 105,
119, 122, 144, 240, 244, 260
- Willems, Dominique 316–317,
319, 333
- Willett, Thomas 3, 18, 147, 163,
188, 212, 260
- Wilmet, Marc 311, 333
- Wiltschko, Martina 210
- Woodcock, Eric Charles
242–243, 249, 256, 260
- Wälchli, Bernhard 240, 260
- Y
- Yllera, Alicia 150, 163
- Z
- Zaefferer, Dietmar 220, 238
- Zaliznjak, Anna 290, 307
- Zamorano Mansilla, Juan Rafael
143
- Zeevat, Henk 191, 212
- Zegrean, Iulia 12, 192, 201, 208
- Zeisler, Bettina 230, 238
- Zimmermann, Malte 186, 193,
196, 199, 202, 212
- Ziv, Yael 179, 183
- Zwarts, Joost 348, 355

Subject index

A

accusative 239, 242–243,
246–247, 251, 258, 305
alethic 41, 43–46, 49–50, 52,
54–55, 57–58, 61
aspect VII, 1, 5–7, 13–18, 38,
45, 47, 62–63, 71, 80, 83–84,
96–98, 100–101, 103–104, 161,
208, 235, 237, 269, 279, 281–
283, 287–291, 296, 298–299,
302–307, 310–312, 327–330,
332–333, 355

B

BNC (British National Corpus)
10, 85, 90, 93, 97, 99–100,
169–173, 240, 287
boundedness 14, 217, 221, 281,
305, 312, 326

C

coercion 6, 17–18, 281–282, 285,
304, 306–307, 312, 332
cognitive linguistics 7, 17–18,
87, 106, 162, 236, 238, 306,
329, 332
collostructional VII, 15, 309–
311, 315–316, 319–320, 326–327,
330, 332
commitment 11–12, 89, 104–
105, 110–111, 114–115, 119–120,
122–124, 136–137, 146, 149,
155, 158–159, 166, 168, 174, 176,
179–180, 182, 186, 188, 194, 213,
216, 223, 225–226, 228–233,
235, 242, 253
conditional 1, 3, 9–10, 45, 50,
62, 65–68, 72–73, 75, 77–78,
83–84, 86, 93–98, 100–104,
106–107, 112, 122, 127, 132–135,
137, 144, 158–159, 209, 213, 216,
295, 319

conjectural 28, 209, 213–214,
216–217, 233, 236
construction grammar 7, 14–15,
18, 166, 259, 306, 309, 314,
329–330
CORLEC 10, 85, 90, 93, 97–102
CREA 84, 97, 101–102

D

deber (*de*) 10, 83–84, 86, 88,
90–91, 96, 98–99, 118, 134, 148
devoir 8–10, 41–64, 86, 105,
148–149, 153, 161–163, 208–209
deure 10–11, 145–146, 148–154,
156, 158–160, 162–163
deictic 2, 7, 92, 94, 208, 214,
218, 222, 227, 229, 256, 339, 347
deontic 43–45, 49–50, 55, 58,
61, 118, 125, 134, 148, 150,
152–153, 156, 160, 177
discourse markers 7, 171, 183,
305
discourse particles 11, 208–209,
211–212

E

East caucasian languages 8, 13,
261–265
see also Andi 261–262,
265–266, 268–274,
276–277, 279
see also Avar 261–277, 279
English VII, 2, 7, 9–10, 12–14,
16–18, 42, 47, 51, 55, 80, 83–87,
89–90, 92, 96–97, 99–101,
103–106, 109, 111–113, 116, 118,
120–124, 126–127, 129–138,
142–144, 148, 161–162, 168–
170, 182–183, 189, 192, 195, 210,
213–214, 216–217, 220, 227,
230–233, 235–240, 242, 247,
252–253, 256–260, 264, 273,

281–288, 300, 303–304, 306,
309, 315, 329–330, 342, 349, 355
enunciative 5, 8, 11, 19–22, 27,
37, 71, 77–78, 165–166, 174–175,
177–178, 180–183, 236
epistemicity 3, 5, 9, 11, 84, 86,
91, 114, 128, 143, 239
epistemic 2–5, 9–12, 16–18,
20, 22–23, 25, 27–29, 34–36,
38, 41–44, 49–50, 56, 61,
64, 83–84, 86–96, 98–101,
103–107, 109–111, 114–120, 122,
124–133, 137, 142–150, 153–163,
165–167, 177, 181, 183, 185–190,
193, 195–196, 199, 209–214,
220, 222, 226, 229–230, 232,
235, 237, 239, 242, 244, 253,
256, 258–259, 279
Estonian 41–43, 46–48, 50–52,
55–57, 59–64
European Portuguese 15, 106,
335–336
event 4–5, 14, 21–24, 29, 42,
45, 52–54, 56–57, 59, 61,
67–68, 71, 84, 92, 100–101,
117, 120, 127, 137, 188, 190–191,
214–215, 218–220, 223,
228–229, 232, 234, 247, 250,
256, 262, 264, 267, 272, 274,
281–283, 285–288, 294–295,
298, 300–302, 305–307, 311–
313, 315, 323–327, 330, 340–343,
345, 347–355
eventuality VII, 1, 3–5, 7, 9–13,
16–18, 47, 83–86, 88–91, 93–97,
99–100, 102–107, 109, 114–115,
118–122, 129–131, 142–150,
153–155, 160–161, 163, 165–166,
181–182, 185, 187–188, 208–213,
235–241, 250, 252, 254–255,
257–262, 264–266, 269, 271,
275–280, 288, 305–306

- see also direct evidentiality
3–4, 20–21, 25, 36, 85–86,
117, 120–122, 143, 147–149,
188–189, 191, 194, 215, 217,
220–221, 230–231, 233, 239,
243–246, 249, 251–254,
256, 259, 294, 316
- see also future evidentiality
3–4, 12–13, 67, 71, 84–86,
89–90, 93–94, 96, 98, 100–
102, 104, 110, 117, 120–122,
143, 147–148, 161, 188, 190,
194, 198, 207, 213–215, 221,
225–227, 229–232, 239–240,
243–245, 249–250, 252,
254, 256, 261–262, 264–267,
269, 271, 273, 275–278
- evidentials 3–5, 7, 10–12,
16–18, 85–87, 93, 104, 119–121,
142–143, 147–148, 150, 154, 161,
181, 185–189, 191, 194, 196–199,
201, 203–204, 206–214, 217,
221, 233, 236–237, 241, 259,
266–267, 279
- evidential expressions 7, 10,
106, 110, 115, 118, 120–121,
124–125, 132, 137–138
- evidential frames 10
- expressive 12, 193, 199–201,
204, 211, 213–215, 219–221, 232,
237, 253
- F
- finite 7, 12–13, 72, 77, 192,
239–240, 242–243, 246,
248–249, 251–252, 255–257,
269, 289, 304
- finiteness 12, 239, 255, 259
- FRANTEXT 72, 216, 234, 310,
328
- French VII, 2, 5, 7–9, 14–15,
19–20, 22, 25–26, 30–37,
41–43, 45, 47–50, 52–53,
55–56, 58, 60, 62, 64–66,
76, 79, 81, 86, 104, 106, 128,
148–149, 153, 163, 216, 227,
238, 242, 281–288, 290–297,
303–305, 309–311, 313–319, 321,
324, 326–333
- future 1–3, 6, 8–10, 15, 17–39,
41–63, 65–67, 70, 77–78,
80–81, 83–84, 86, 88–89,
91–94, 96, 100–101, 103–106,
109, 111–113, 117, 123, 126–127,
132–137, 143, 146, 149–150, 152,
156, 158–160, 163, 213–214, 234,
243, 256–258, 270, 276, 278,
313, 315–316, 328
- see also future perfect 10,
83–84, 92–94, 96, 100–101,
104, 276
- futur antérieur* 8, 19, 22, 37
- futur simple* 8, 19, 22, 37–38
- G
- German 18–20, 22, 24–26,
28–38, 86, 105, 153, 162,
185–189, 191–192, 194, 201, 207,
209–211, 242, 259, 271, 313
- grammaticalization 4, 6, 8–9,
13, 16, 42–43, 57, 61–63, 65,
68–70, 77, 80, 104–105, 142,
145, 150, 153, 160, 163, 182–183,
185, 258, 261, 264–265, 279
- H
- haber (de)* 10, 83–84, 96, 98–99
- haver de* 11, 145–146, 149–150,
152–154, 156, 158–160
- Hungarian VII, 14, 281–283,
287–288, 290, 296–298,
300–306
- I
- imminence 9, 46, 65–66,
68–70, 72, 77–78, 80
- imparfait* 5, 8, 14–15, 66, 81,
284, 286, 296, 309–310,
328–331, 333
- imperfective 5–6, 14–15, 43, 84,
281–282, 284–285, 287–288,
294, 298, 300, 303–305, 311,
319, 322–326
- imperfect 9, 65–69, 71–78, 86,
268, 276, 278, 286–287, 304,
315, 321–322, 326
- inference 3–4, 11, 22, 28, 70,
83, 86–88, 93, 103, 119, 145,
148–155, 158–160, 188, 190–191,
194–196, 214, 230, 243, 248,
252, 254, 256, 265
- inferential VII, 4, 10–11, 15, 28,
83–90, 92–93, 96–101, 103–
104, 138, 145–146, 148–150,
152–155, 158–160, 185, 187–188,
194–196, 202, 207, 211, 214,
216–217, 220, 239, 241, 248,
250, 252, 259, 264, 310, 314
- interrogative 160, 190, 192,
213–216, 220–221, 224, 229,
237, 269
- irrealis* 10, 83–84, 86, 93–96,
103–104, 111, 124, 126, 227, 319
- L
- Latin 69, 150, 239, 242, 249,
251, 253, 256–260
- Lithuanian VII, 8, 12–13, 144,
239–241, 244–247, 249–253,
255–260
- M
- mirativity 4, 7, 12, 17–18, 125,
213–214, 220–221, 229–233,
236
- modality VII, 1–5, 8–11, 16–20,
30, 32, 36–38, 41–45, 55–58,
61–62, 64, 80, 83, 86–87, 89,
91–92, 97–99, 103–107, 109,
114–116, 125–127, 129–131,
143–149, 153–156, 159–163,
165–166, 182, 185, 187–188, 191,
209–212, 227, 229, 235, 237,
259–260, 279, 288, 305–306,
328
- see also epistemic modality
3–5, 10–11, 17, 38, 43, 56,
83, 87, 89, 91, 104, 106–107,
109, 114–116, 125–127,
129–131, 144–149, 153–156,
160–163, 165–166, 185,
187–188, 209, 211, 235, 237,
259, 279
- modal auxiliaries 2, 101, 105,
116, 118, 125, 143, 227, 232
- modal verbs 1, 7–11, 16, 37,
41–43, 48, 50, 56–57, 60–62,
86, 103–104, 145–150, 153, 155,
160, 162–163
- motion verbs 8, 15, 335–338,
346–348, 350–351, 353–354

- must* 10–11, 41, 44, 46–47, 49, 53–56, 58–60, 83–84, 87–91, 93, 96, 98–100, 118, 145–152, 154–155, 188, 190, 195, 197–198, 219, 338
- N**
- newspapers 14, 19–20, 30–31, 33, 36–37, 97, 102, 113, 318, 324
- non-finite 7, 12–13, 239–240, 242–243, 246, 248–249, 255–257, 269, 289, 304
- O**
- objective 1–2, 5, 9, 11, 15, 41, 44–46, 49–54, 56–58, 60–61, 65–69, 71, 78, 90, 109, 127, 150–151, 153, 159, 182, 353
- Old Catalan 10–11, 145–146, 150, 152–153, 156, 159–160, 163
- P**
- participle 12–13, 62, 239–243, 246–247, 249–250, 254–255, 257–258, 269, 278, 283
- passé composé* 8, 14, 71, 281–283, 285, 305, 307, 309–310, 328–331, 333
- path scale 15, 335, 348, 350–351, 353
- perception 2, 4–5, 13, 75–76, 85–86, 118, 175, 230, 237, 239, 241–245, 247–248, 250, 252, 254, 256–259, 311, 323, 331
- perfective 5–7, 14–15, 17, 58, 258, 265, 281–288, 300, 303–305, 311–312, 319, 321–325
- perfectivity 14, 281, 283, 286–287, 291, 302
- periphrastic 2, 8–9, 31, 37, 45, 53, 65–66, 69, 77, 127, 269, 319
- pidama* 8–9, 41–43, 46–48, 50–61, 63–64
- poder* 10–11, 118, 145–146, 148–151, 155–160, 163
- postmodal 9, 41–42, 45–46, 48, 60–61
- possibility 21, 23, 25, 42, 48, 50, 89, 116, 128, 145, 155–160, 177, 180, 189–190, 195, 198, 201, 203, 298, 311, 342, 349, 353
- prefixation 281, 288, 297, 299, 302
- prepositions 15, 31, 318, 335–338, 342–343, 348, 353
- present 1–3, 7–10, 17–21, 27–29, 44–45, 47–49, 54, 56–62, 65–68, 71–73, 75, 77–78, 87, 89, 91–93, 98, 105, 109, 111, 114, 126, 129–130, 133–135, 137, 159, 165–167, 169, 181, 210, 213, 215–216, 223–224, 232, 239, 241, 243, 249, 256, 258, 261, 263–268, 270, 274, 278, 282–285, 298, 304, 313, 315, 317, 324, 336, 340
- prospective 9, 23–25, 35, 43, 47, 65–69, 71–78
- Q**
- questions 7, 12–13, 49, 76–77, 106, 163, 167, 189–190, 192–193, 198, 202–203, 207, 209–210, 213–237, 273, 276, 310–311, 325
see also conjectural questions 28, 209, 213–214, 216–217, 233, 236
see also unresolvable questions 12, 213–221, 231–234
see also rhetorical questions 12, 21, 25, 27, 29, 168, 191, 213–218, 221–235, 237, 325
- R**
- reportative 10, 83–86, 90–91, 94–104, 120–121, 197, 203, 208, 276, 278
- reporting 109–110, 121–122, 124–125, 132, 136–137, 144, 178
- resultative 13, 93, 96, 99, 103, 261–267, 269, 271–273, 277–279, 283, 298, 311
- Russian 110, 242, 248, 252, 258, 268, 279–284, 287–291, 293–298, 300, 302–307
- S**
- semantics 6, 10, 13, 15, 17–18, 21, 94, 105–107, 143–145, 148, 160, 207–212, 218, 236–237, 248–249, 252, 254, 259, 262, 268–269, 275, 278–279, 298, 300, 304–307, 310–313, 316, 320, 326–327, 329–330, 332, 335–336, 348, 353–355
- scalar 15, 94, 218, 306, 335–336, 348, 350–351, 353–355
- sollen* 3, 18, 153, 162
- Spanish 66, 78, 83–86, 88–89, 91–93, 96–104, 107, 109, 111–113, 116–118, 120–128, 130–144, 148, 161, 208, 257
- speech acts 12, 125, 193, 209, 211–215, 217, 220–221, 225, 229–231
- subjective 1–2, 5, 9, 11, 18, 41, 44–46, 49–51, 53–58, 60–61, 65–69, 71–72, 75–78, 90–91, 103, 118, 137, 145–146, 151, 153, 157, 160, 174–177, 181–182, 259
- subjectification 145–146, 153, 160, 162–163
- synthetic 9, 49, 65, 77–78, 126, 319
- T**
- Tagalog 185–191, 194, 207–208, 210–212
- telicity 281, 299, 301, 305–306, 323, 335–336, 348–351, 354
- temporal use 8–9, 19–20, 23–24, 27, 30, 32, 36–37, 49, 61
- tense VII, 1–3, 5–9, 14–39, 41, 44–45, 49, 51, 58–63, 68, 80, 83–84, 86, 88, 91–92, 96–98, 100–101, 103–104, 111–112, 126–127, 137, 143, 146, 158, 161, 208, 237, 243, 249, 256, 264, 266, 268–270, 273, 278–279, 282–283, 285–288, 294, 296, 300, 304–307, 309–317, 319–321, 323–330, 355
- time 1–2, 5, 9, 16–18, 21, 23–24, 26–27, 29, 31, 41–43, 48, 57, 59–60, 63, 67, 71, 76–78, 84, 86–88, 91–93, 104–106, 109–110, 112, 119, 124, 126–128, 135, 137, 140–141, 172, 180, 203, 218, 225, 229, 250, 255–257, 262–264, 266–267, 275–276, 282–284, 291, 294, 305–307, 310–311, 331, 340–341, 343, 345, 347, 351–354

U

ulteriority 1, 3, 9, 65–69, 71–72,
75, 77–78

V

verb classes 281, 301, 355

W

will 2, 4, 10–11, 19–20, 22–24,
26–28, 30, 35–36, 42, 48, 50,
54, 66–67, 71, 83–84, 86–93,
96, 100–101, 104, 109–112,
114–118, 121–129, 131–141,
146–147, 149–151, 155, 158, 160,
166, 169, 173, 175–176, 185–186,
188, 191, 193–196, 198–199,
201–205, 207, 217, 228,
242–244, 249, 251, 255–256,

261–262, 282–284, 288, 290,
296, 299, 310–311, 317–318, 321,
327, 336–338, 340, 346

wohl 11–12, 185–189, 191–207,
209–212

wollen 3–4, 26

Y

yata 11–12, 185–192, 194–199,
202–207

After an introductory chapter that provides an overview to theoretical issues in tense, aspect, modality and evidentiality, this volume presents a variety of original contributions that are firmly empirically-grounded based on elicited or corpus data, while adopting different theoretical frameworks. Thus, some chapters rely on large diachronic corpora and provide new qualitative insight on the evolution of TAM systems through quantitative methods, while others carry out a collostructional analysis of past-tensed verbs using inferential statistics to explore the lexical grammar of verbs. A common goal is to uncover semantic regularities and variation in the TAM systems of the languages under study by taking a close look at context. Such a fine-grained approach contributes to our understanding of the TAM systems from a typological perspective. The focus on well-known Indo-European languages (e.g. French, German, English, Spanish) and also on less commonly studied languages (e.g. Hungarian, Estonian, Avar, Andi, Tagalog) provides a valuable cross-linguistic perspective.

ISBN 978 90 272 0096 9



JOHN BENJAMINS PUBLISHING COMPANY