

Time in Languages, Languages in Time

EDITED BY

Anna Čermáková

Thomas Egan

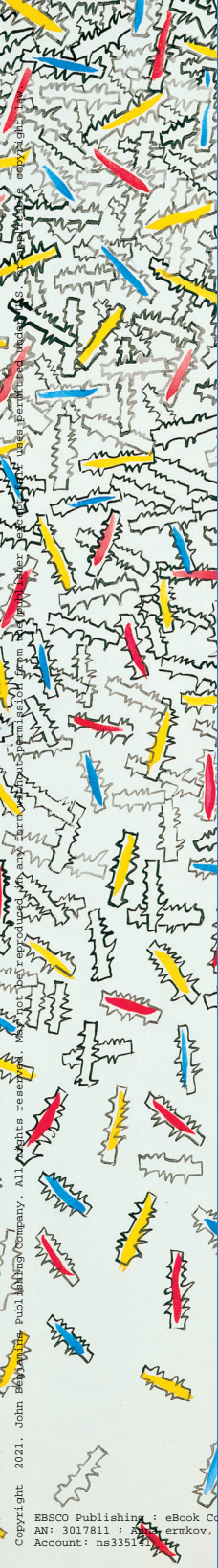
Hilde Hasselgård

Sylvi Rørvik

Studies in Corpus Linguistics

101

JOHN BENJAMINS PUBLISHING COMPANY



Time in Languages, Languages in Time

Studies in Corpus Linguistics (SCL)

ISSN 1388-0373

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

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

General Editor

Ute Römer
Georgia State University

Founding Editor

Elena Tognini-Bonelli
The Tuscan Word Centre/University of Siena

Advisory Board

Laurence Anthony
Waseda University

Antti Arppe
University of Alberta

Michael Barlow
University of Auckland

Monika Bednarek
University of Sydney

Tony Berber Sardinha
Catholic University of São Paulo

Douglas Biber
Northern Arizona University

Marina Bondi
University of Modena and Reggio Emilia

Jonathan Culpeper
Lancaster University

Sylviane Granger
University of Louvain

Stefan Th. Gries
University of California, Santa Barbara

Susan Hunston
University of Birmingham

Michaela Mahlberg
University of Birmingham

Anna Mauranen
University of Helsinki

Andrea Sand
University of Trier

Benedikt Szmrecsanyi
Catholic University of Leuven

Elena Tognini-Bonelli
The Tuscan Word Centre/University of Siena

Yukio Tono
Tokyo University of Foreign Studies

Martin Warren
The Hong Kong Polytechnic University

Stefanie Wulff
University of Florida

Volume 101

Time in Languages, Languages in Time

Edited by Anna Čermáková, Thomas Egan, Hilde Hasselgård and Sylvi Rørvik

Time in Languages, Languages in Time

Edited by

Anna Čermáková

Charles University, Prague

Thomas Egan

Inland Norway University of Applied Sciences

Hilde Hasselgård

University of Oslo

Sylvi Rørvik

Inland Norway University of Applied Sciences

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.

Cover design: Françoise Berserik
Cover illustration from original painting *Random Order*
by Lorenzo Pezzatini, Florence, 1996.

DOI 10.1075/scl.101

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2021024183 (PRINT) / 2021024184 (E-BOOK)

ISBN 978 90 272 0968 9 (HB)
ISBN 978 90 272 5896 0 (E-BOOK)

© 2021 – 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

CHAPTER 1	
Time in languages, languages in time	1
<i>Anna Čermáková, Thomas Egan, Hilde Hasselgård & Sylvi Rørvik</i>	
CHAPTER 2	
Slavery and Britain in the 19th century	9
<i>Tony McEnery, Helen Baker & Vaclav Brezina</i>	
CHAPTER 3	
Diachronic change in the ordering of kinship binomials: A contrastive perspective	39
<i>Anna Čermáková</i>	
CHAPTER 4	
Tracing processes in auxiliarization – Time sufficiency verbs from a Norwegian-Swedish-English contrastive perspective	67
<i>Mats Johansson & Lene Nordrum</i>	
CHAPTER 5	
Translating verbal tenses between tensed and tenseless languages: A contrastive study of multilingual corpora	95
<i>Cristina Grisot & Juan Sun</i>	
CHAPTER 6	
From language to language, from time to time: Echoic binomials from an English-German-Swedish perspective	129
<i>Magnus Levin & Jenny Ström Herold</i>	
CHAPTER 7	
<i>Just a moment</i> – Brief times in English and Norwegian	155
<i>Thomas Egan & Siri Füst Skogmo</i>	
CHAPTER 8	
<i>Then and now</i> in English and French: Parallel patterns?	181
<i>Diana M. Lewis</i>	

CHAPTER 9

Time adverbials in English and Norwegian news discourse 201

Hilde Hasselgård

CHAPTER 10

Minutes of action! A contrastive analysis of time expressions
in English and Norwegian football match reports 229*Signe Oksefjell Ebeling*

CHAPTER 11

Cross-disciplinary and cross-linguistic uses of tensed verb phrases
in the methods sections of master's theses 255*Sylvi Rørvik*

CHAPTER 12

The expression of time in English and Czech children's literature:
A contrastive phraseological perspective 283*Markéta Malá, Denisa Šebestová & Jiří Milička*

Index 305

CHAPTER 1

Time in languages, languages in time

Anna Čermáková, Thomas Egan, Hilde Hasselgård &
Sylvi Rørvik

Charles University / Inland Norway University of Applied Sciences /
University of Oslo

1. Introduction

*Of all the scientific intangibles
that shape our lives,
time is arguably the most elusive
– and the most powerful.*

(Langone 2000: 7, cited in Evans 2003)

Time, together with place, is one of the fundamental dimensions by which we define our ‘space’: linguistically, historically, socially, and culturally. As Evans (2004: 3) argues, “it seems almost impossible to conceive of what our world of experience might be like in the absence of time”. Time appears to be a “natural” concept, as if “lying outside of society and beyond human control” (Warf 2008: 2) – recent scholarship across disciplines, however, reveals that, despite its seeming objectivity, it is very much a social construction and “every society develops different ways of dealing with and perceiving” time (ibid.). The topic of time and the various means languages have and use to express various time relations and concepts have always fascinated linguists. An important area of research covers, for example, metaphors (e.g. Lakoff & Johnson 1980, 1999), and more broadly, what Evans (2004: 5) calls “the linguistic problem of time”, i.e. “why do we use language pertaining to motion through three-dimensional space [...] in order to think and talk about time?”

Arguably, time is primarily encoded lexico-grammatically through verbs – however, tense and aspect systems vary enormously across languages (see, e.g., Ayoun et al. 2018; Ebeling & Hasselgård 2015). Time can furthermore be encoded as part of the lexical meaning of verbs. But other word classes, too, express time and temporal relations, notably adverbs and prepositions. The word *time* itself is

one of the most frequent nouns in English and its sheer frequency means it is part of numerous patterns, phrases, and constructions expressing a variety of temporal meanings. Linguistic expressions of time contribute to structuring our experience of the world by locating events on a time scale. In addition, temporality may be part of discourse organization, representing not only the temporal order of events but also the internal, consecutive, ordering of arguments or other portions of a text (Halliday & Hasan 1976: 263).

One way of looking at time in language, is therefore, through examining language structures, patterns and lexis expressing and encoding time. However, language itself is substantially influenced by time. Language change throughout time is one of its inherent characteristics – “all synchronic states are the result of a long chain of diachronic developments, the construction of complete explanations for linguistic structures requires attention to the diachronic dimension” (Bybee 2010: 945).

The cross-linguistic study of time expressions is fascinating because it can give “important insights into the conceptualisation of time” (Evans 2004: 5), but it is also interesting in itself. Although “all natural languages we know of have developed a rich repertoire of means to express temporality” (Klein 1994: 1), it is well known that languages vary substantially in the means of expressing time that they have available to them. The difference may be grammatical, such as the lack of grammatical tense in Chinese and of a grammaticalized progressive aspect in German and the Scandinavian languages. Or it may be lexical, as seen in Johansson’s (2007: 46) case study of the Norwegian word *døgn* (meaning the 24 hours between one midnight and the next) and its English, German and French correspondences in the absence of an equivalent word. It may also be that two or more languages possess similar inventories of lexico-grammatical expressions of time, but select differently from them in certain situations, or that superficially similar lexico-grammatical categories express different temporal meanings, as in the case of the French ‘passé composé’ and the German ‘Perfekt’ on the one hand, and the English present perfect on the other (de Swart 2007).

Furthermore, languages and cultures may divide up the semantic space of ‘time’ differently. One example is lexical expressions of times of the day: according to Johansson (2007: 42), the English *morning* ends later than its Norwegian cognate *morgen*. Moreover, approximate references to times and periods may have different meanings: how long does an English ‘moment’ last in comparison with corresponding expressions in other languages? How come two weeks may be referred to as fourteen days in Norwegian (*fjorten dager*) but as fifteen (*quinze jours*) in French? And how do translators cope with such apparent mismatches between systems and preferences?

The present volume brings together a collection of articles on languages and time, most of which build upon papers presented at the ICAME40 preconference contrastive workshop, which took place in Neuchâtel in 2019 and was convened

by the editors of this volume. The theme of the workshop was ‘Time in Languages, Languages in Time’. The contents of the book reflect these dual perspectives, with Part 1 containing three papers dealing with aspects of language change over time, and Parts 2 and 3 containing eight papers contrasting various temporal expressions in two or more languages. The call for papers encouraged a phraseological perspective on temporal expressions, and this is visible in several of the chapters which deal with multi-word units of various kinds: collocations, colligations, binomials and n-grams.

All the contributions are based on or driven by multilingual corpus data from both parallel and comparable corpora. The languages represented include English in comparison with one or more of the following: Czech, French, German, Mandarin, Norwegian and Swedish. The parallel corpora, containing original and translated texts as they do, allow the study of translation practices as well as the systematic cross-linguistic comparison of words, phrases and grammatical features. This is apparent in Levin and Ström-Herold’s chapter, for example, which is concerned with translation correspondences across the three languages English, German and Swedish but also pays due attention to the comparison of binomials that are found in originals in all three languages. Similarly, Grisot and Sun’s study of the translation of verb phrases between languages with and without tense marking shows how the examination of translation correspondences “may serve as a guide to the interpretation of meaning” (Johansson 2007: 30). The studies based on comparable corpora, on the other hand, need to establish the basis for comparison on something other than translation correspondence, such as frequency data and lexical similarity (as in Ebeling’s chapter), syntactic function (as in Hasselgård’s chapter), or extraction methods, as in the chapter by Malá et al. This paper also explicitly discusses the problem of identifying comparable multi-word units in typologically distant languages and of the potential contribution of n-gram-based approaches to contrastive linguistic studies (see also Granger 2014; Čermáková & Chlumská 2017).

The chapters display a range of methodological approaches, both corpus-based and corpus-driven. Several methodological challenges are addressed and solutions proposed. The results of the various studies testify to the existence of robust systematic correspondences between languages in certain cases, and the lack of such correspondences in others. They also testify to the pervasiveness and great diversity of temporal expressions across time, languages and genres and to the many referential and pragmatic functions that they serve.

2. Languages in time

Part 1 of the book, ‘Languages in time’, opens with an interdisciplinary chapter, grounded in both corpus linguistics and history research, by Tony McEnery, Helen

Baker and Václav Brezina. In a study that has grown even more topical since it was first proposed, they approach the issue of slavery in Britain in the 19th century through a corpus-based investigation of historical news articles, posing the question of how to explore concepts and their change over time in large corpora. Their approach involves a novel method called ‘Usage Fluctuation Analysis’ (McEnery et al. 2019), which uses collocation to locate diachronic shifts in the discussion of the concept. They also address the important question of how to deal with historical data of poor (technical) quality.

Next, Anna Čermáková takes a diachronic and contrastive perspective on binomials involving kinship terms in English and Czech children’s literature. Discussing the gradual diachronic reversal of the male → female order in binomial kinship terms, and more specifically the phrase *father and mother*, Čermáková proposes that this change is linked to a more general development in the discourse, namely the shift towards greater informality.

Mats Johansson and Lene Nordrum report on a study entitled ‘Tracing processes in auxiliarization – time-sufficiency verbs from a Norwegian-Swedish-English contrastive perspective’. They assume that participant-external sufficiency, as exhibited by the semi-auxiliary *rekke*, is a modal category, which can be described in terms of a semantic map of modality (van der Auwera & Plungian 1998). A central proposal is that *rekke* is undergoing grammaticalization from a concrete spatial verb (‘reach’) towards one that contains a component of time-sufficiency and participant-internal possibility. The grammaticalization process involves a relation between semantic and syntactic development that makes *rekke* a (semi-)auxiliary with multi-layered modal interpretation.

3. Expressing and translating time

Part 2, ‘Expressing and translating time’, comprises four chapters. Cristina Grisot and Juan Sun’s paper, ‘Translating verbal tenses between tensed and tenseless languages: a contrastive study of multilingual corpora’, contrasts languages that have markedly different ways of grammatically encoding time through tense and aspect. That is, they study translations of English and French verbal tenses into Mandarin Chinese and conversely, Mandarin verb phrases translated into English and French. They find, for example, that English and French present-time verbal tenses are most frequently translated by linguistically non-marked VPs in Mandarin, past-time verbal tenses by aspectual, temporal markers and linguistically non-marked VPs, and future-time verbal tenses by temporal markers.

Magnus Levin and Jenny Ström Herold’s chapter is entitled ‘From language to language, from time to time: echoic binomials from an English-German-Swedish

perspective'. This data-driven study finds that echoic time binomials are equally common in English and German but much more frequent in Swedish originals, largely due to the frequency of Noun-Preposition-Noun constructions such as *timme efter timme* ('hour after hour'). When translations of time binomials in their material are not echoic, meanings still tend to be expressed by related recurring phraseological patterns, e.g., *line by line* > *en rad i taget* ('a line at a time'); i.e., recurrent temporal meanings tend to be expressed by recurrent patterns.

Thomas Egan and Siri Først Skogmo investigate the coding of points in time and brief intervals of time by English *moment* and its closest Norwegian correspondence *øyeblikk* ('wink of an eye') in their chapter 'Just a moment – brief times in English and Norwegian'. Both lexemes are commonly found to encode time intervals, but *moment* is almost twice as likely as *øyeblikk* to refer to a point in time. According to the authors, this suggests that *moment* is somewhat more polysemous than its Norwegian counterpart. The translations corroborate this conclusion: *øyeblikk* is more likely to be translated as *moment* than vice versa.

Using two comparable corpora of present-day European French and British English political speeches and discussions, Diana M. Lewis investigates '*Then* and *now* in English and French: parallel patterns?' *Then*, *now*, and their French counterparts *alors* and *maintenant*, exemplify a cross-linguistic tendency for temporal adverbs to grammaticalize into markers of rhetorical relations, although all are used in both temporal and argumentative senses. The adverbs are more frequent in the discussions than in the speeches in both languages, but the genre difference is greater in French. In spite of great cross-linguistic similarities, the French and English adverbs seem to be at different stages of grammaticalization, with *maintenant* being less frequent and less grammaticalized than *now*. By contrast, *alors* occurs in a wider range of rhetorical contexts, and looks more bleached and grammaticalized than *then*.

4. Time in languages and genres

Part 3 of the volume, 'Time in languages and genres' comprises four chapters, each of which focuses on language comparison within a specific genre. First in this section, Hilde Hasselgård studies time adverbials in English and Norwegian news discourse. The comparison concerns the syntactic realizations, semantic types and positions of the adverbials. It also includes the lexical realizations of time adverbials in both languages, partly through the lens of lexical priming (Hoey 2005). The distribution of syntactic and semantic types of time adverbials is similar between the languages, but there are differences in adverbial placement, particularly in clause-medial position, where English is more restrictive than Norwegian. The

lexical comparison shows that the languages sometimes differ in the ways similar meanings are realized. Furthermore, the lexical priming of some frequent lexical items reveals language-specific, lexeme-specific and possibly register-specific patterns.

A highly specialized register is examined in Signe Oksefjell Ebeling's chapter 'Minutes of action! A contrastive analysis of time expressions in English and Norwegian football match reports'. The cognates *minutes* and *minutter*, which are very frequent in this register, are used as a starting point for identifying patterns by means of n-grams and collocations. Zooming in on the patterns *on # minutes* and *etter # minutter* ('after # minutes'), Ebeling finds that the English expression clearly prefers final position while the Norwegian expression has a slight preference for initial position. Furthermore, English has a clear preference for the past tense in clauses containing this expression, while Norwegian also makes some use of the (historic) present tense. However, both patterns are typically associated with the achievements of players, often involving a goal scored.

Sylvi Rørvik studies tense usage in academic writing in the chapter 'Cross-disciplinary and cross-linguistic uses of tensed verb phrases in the methods sections of master's theses'. Her corpus consists of the methods sections of 150 master's theses in English and Norwegian within the three disciplines of chemistry, linguistics, and sociology, all submitted at Norwegian universities. The cross-disciplinary differences are found to be greater than the cross-linguistic ones: in both languages, present tense verb phrases are most frequent in linguistics, less frequent in sociology, and least frequent in chemistry. However, the cross-disciplinary differences are greater in Norwegian than in English, which may indicate that those who write in Norwegian are better at tailoring their use of tense to the conventions of their discipline.

In their study 'The expression of time in English and Czech children's literature: A contrastive phraseological perspective' Markéta Malá, Denisa Šebestová and Jiří Milička are facing a cross-linguistic methodological challenge: how to identify comparable multi-word units in typologically distant languages, in their case Czech (a highly inflectional language with a relatively free word order) and English (a predominantly analytical language with a more fixed word order). It is suggested that this challenge can be at least partially overcome with the help of the tool Engrammer (<http://milicka.cz/en/engrammer/>), which enables the exploration of unordered n-grams (i.e. n-grams with positional mobility). A good number of idiomatic (i.e. relatively fixed) time expressions are noted in both languages, but Czech employs a wider range of these – often stylistically marked ones – than English. The results indicate that time plays an important role in structuring the text in children's fiction in both languages. Even though the formal means of expressing time may differ between the languages, register appears to influence substantially the way time is framed in children's literature.

Being “[deeply] rooted in the structural organisation of language” (Klein 1994: 1), time is encountered across the linguistic system: lexically, phraseologically, morpho-syntactically and linearly. The present volume shows that languages differ in the ways in which they employ the various resources for expressing time. The differences may be systemic, e.g. regarding whether or not aspect and tense are grammaticalized in a language, or conventional, e.g. regarding the frequency and collocational patterns of (extended) lexical items referring to time, or textual norms for the choice of tense forms in specific contexts. Moreover, the passing of time changes the way in which certain concepts are interpreted – changes which may not follow the same path in different languages. The contrastive perspective of this volume corroborates the claim that multilingual corpora can “increase our knowledge of language-specific, typological and cultural differences, as well as of universal features” (Aijmer & Altenberg 1996: 12). The language comparison throws “special features of the languages compared into relief, including preferred ways of expressing similar meanings” (Johansson 2012: 64). It is hoped that the following chapters will not only provide insight into the topics studied but also serve as a source of inspiration for further explorations of the fundamental category of time across registers, languages and time.

In memoriam

Mats Johansson sadly passed away before this book was published. Not only did he contribute fully to Chapter 4, he also gave constructive input to many of the other authors at the workshop which gave rise to this volume. He will be sorely missed by colleagues in the field of contrastive corpus linguistics.

References

- Aijmer, Karin & Altenberg, Bengt. 1996. Introduction. In *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*, Karin Aijmer, Bengt Altenberg & Mats Johansson (eds), 11–17. Lund: Lund University Press.
- Ayoun, Dalila, Celle, Agnès & Lansari, Laure. 2018. *Tense, Aspect, Modality, and Evidentiality: Crosslinguistic Perspectives*. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.197>
- Bybee, Joan. 2010. Diachronic Linguistics. In *The Oxford Handbook of Cognitive Linguistics*, Dirk Geeraerts & Hubert Cuyckens (eds), 945–986. Oxford: OUP.
- Čermáková, Anna & Chlumská, Lucie. 2017. Expressing PLACE in children's literature. Testing the limits of the n-gram method in contrastive linguistics. In *Crosslinguistic Correspondences: From Lexis to Genre*, Hildegunn Dirdal & Thomas Egan (eds), 75–95. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.191.03cer>

- de Swart, Henriëtte. 2007. A cross-linguistic discourse analysis of the Perfect. *Journal of Pragmatics* 39: 2273–2307. <https://doi.org/10.1016/j.pragma.2006.11.006>
- Ebeling Signe Oksefjell & Hasselgård, Hilde (eds). 2015. *Cross-linguistic Perspectives on Verb Constructions*. Newcastle upon Tyne: Cambridge Scholars.
- Evans, Vyvyan. 2004. *The Structure of Time: Language, Meaning and Temporal Cognition* [Human Cognitive Processing 12]. Amsterdam: John Benjamins. <https://doi.org/10.1075/hcp.12>
- Granger, Sylviane. 2014. A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast* 14(1): 58–72. <https://doi.org/10.1075/lic.14.1.04gra>
- Halliday, Michael A. K. & Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.
- Hoey, Michael. 2005. *Lexical Priming. A new Theory of Words and Language*. London: Routledge.
- Johansson, Stig. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies* [Studies in Corpus Linguistics 26]. Amsterdam: John Benjamins.
<https://doi.org/10.1075/scl.26>
- Johansson, Stig. 2012. Cross-linguistic perspectives. In *English Corpus Linguistics: Crossing Paths*, Merja Kytö (ed.), 45–68. Amsterdam: Rodopi.
- Klein, Wolfgang. 1994. *Time in Language*. London: Routledge.
- Lakoff, George & Johnson, Mark. 1980. *Metaphors we Live by*. Chicago IL: The Chicago University Press.
- Lakoff, George & Johnson, Mark. 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York NY: Basic Books.
- Langone, John. 2000. *The Mystery of Time: Humanity's Quest for Order and Measure*. Washington DC: National Geographic.
- McEnery, Tony, Brezina, Vaclav & Baker, Helen. 2019. Usage Fluctuation Analysis. A new way of analysing shifts in historical discourse. *International Journal of Corpus Linguistics* 24(4): 413–444. <https://doi.org/10.1075/ijcl.18096.mce>
- van der Auwera, Johan & Plungian, Vladimir A. 1998. Modality's semantic map. *Linguistic Typology* 2: 79–124. <https://doi.org/10.1515/lity.1998.2.1.79>
- Warf, Barney. 2008. *Time-Space Compression: Historical Geographies*. London: Routledge.
<https://doi.org/10.4324/9780203938058>

Slavery and Britain in the 19th century

Tony McEnery, Helen Baker & Vaclav Brezina

Lancaster University

This study uses a corpus of just under two billion words from one historic British newspaper, the *Liverpool Mercury*, to explore shifting attitudes to slavery in Britain in the nineteenth century in the context of a port city that benefitted from the trade. In doing so, we explore three methodological issues – how to explore concepts in large corpora, how to do this over time and how to deal with poor quality data. Our approach to the study of concepts through time uses a new approach to looking at word usage change over time, Usage Fluctuation Analysis (McEnery, Brezina & Baker 2019). Our exploration of the issue of poor quality data is motivated by the variable quality of the OCR texts which constitute our nineteenth-century newspaper corpus data. Problems in data quality bedevil work on large-scale text collections of historic material. In this paper we will show that collocation, the core technique of UFA, can be used on such data if appropriate settings are chosen that minimise the problems arising from poor quality electronic text, permitting the exploration of corpus data at scale.

1. Introduction

In this chapter, we discuss changes in the representation of slavery in nineteenth-century Britain through the lens of texts published in one prominent local newspaper. In doing this, we will deal with three methodological issues in pursuit of a single research question. The methodological issues relate to data quality and the scale of the dataset. The research question is how, during the nineteenth century, did the representation of slaves and slavery vary diachronically in one British newspaper? The newspaper in question, the *Liverpool Mercury*, was chosen because it was a newspaper published in and for one of Britain's principal trading ports, sitting at one corner of the so-called 'golden triangle' which drove the transatlantic slave trade. Importantly, through the trade in cotton and manufactured goods with the United States, Liverpool retained strong economic links to a slave-owning economy for most of the century. So, while emancipation of slaves occurred in the early part

of the century throughout the British Empire, Liverpool retained a strong economic interest in slavery, leading some (e.g. Sherwood & Sherwood 2007: 1) to claim that Britain's involvement with slavery extended well beyond the 1833 Slavery Abolition Act. A good example of such involvement comes from Liverpool businesses, like Fraser, Trenholm and Company, which had personal and financial links to the slave trade that meant that its prosperity was "heavily staked on a Confederate victory" (Loy 1997: 349) and the continuation of slave ownership in the American Civil War.

But to explore how slaves and slavery are discussed in the *Liverpool Mercury*, we need to address the methodological issues raised. Accordingly, this chapter is divided into two parts. In the first part (Sections 2 and 3), we consider (i) how to deal with data which contains a considerable degree of error, (ii) how to do so in a corpus which is far too large to explore through qualitative approaches alone – our *Liverpool Mercury Corpus* contains 1,979,108,742 words, from 1811, when the newspaper was founded, to 1900 and (iii) how to analyse changing patterns of usage over time in large scale datasets through a specific technique, Usage Fluctuation Analysis (UFA). Key to the discussion in this first part is the extent to which we can reliably extract collocates from data containing errors as collocation is central to the analysis in the second part of the chapter and underpins the UFA analysis. In the second part of the chapter, we establish the historical framework within which our results may be interpreted (Section 4) and briefly describe the *Liverpool Mercury* in more detail (Section 5). We then proceed, through a process of triangulation, both to outline attitudes to slavery in the *Liverpool Mercury* and to explore the degree to which UFAs provide a helpful way of dealing with such questions in large, noisy, datasets (Sections 6 and 7). The approach to analysis taken in Section 6 is a discourse historical approach (Wodak 2001) to corpus assisted discourse studies (Partington 2003; Baker 2006; Partington et al. 2013). The discourse historical approach requires us to situate our analysis of discourse in historical context, while the corpus assisted discourse analysis helps us in "uncovering, in the discourse type under study ... what we might call non-obvious meaning, which might not be readily available to naked eye perusal" (Partington et al. 2013: 11). By triangulating between the historical context, corpus assisted discourse analysis and the UFA, we show how the approaches can inform the analysis by providing mutual challenge, support and an enhanced explanation of the patterns revealed by techniques such as UFA.

2. Data and data quality

The data used for this study has been converted into machine readable text by the British Library using optical character recognition (OCR), as discussed in McEnery, Baker and Dayrell (2019). The text itself was originally intended to be used to guide readers to photographic images of pages on which the text might be found rather

than as the basis of a corpus per se. We were given access to the OCR text for the whole of the nineteenth century for the *Liverpool Mercury* and a number of other newspapers. On the basis of that text, we constructed corpora for each newspaper. Collectively we call these the *British Library Nineteenth-century Newspaper Corpus* (BLNCNC). While the challenges of working with poor quality data running to billions of words are substantial, the issue we want to focus on here is fundamental: the quality of the OCR data itself. A casual inspection of the data will show that the text is far from perfect – indeed the poor quality of such data has been identified as “a major issue” (Prescott 2014: 336) and has been characterised as an intractable problem for the use of such data (Hitchcock 2013). For corpus linguists, therefore, a key question has to be to what extent corpora based on such OCRed material are usable. If the effect of OCR errors is substantial for corpus analysis, and the issue as intractable as Hitchcock suggests, then using such a resource for corpus analysis would be impossible. Previous research has shown that OCR quality in the BLNCNC is variable and poor – according to Tanner et al. (2009), on average 78% of word forms are correctly identified in the corpus, though this figure falls to 68.4% if content words alone are considered. So OCR is a real issue in the corpus.

To address this issue, we relied on the work of Joulain (2017) as her painstaking close reading and experimental analysis of the effect of OCR errors on corpus analysis provide us with the warrant to conduct our analysis on noisy data. Joulain was able to establish a level of error in the BLNCNC by identifying the subset of the BLNCNC which was present in a hand-corrected set of nineteenth-century newspaper articles. This hand-corrected data, which acted as a gold standard, was the *Corpus of Nineteenth-century Newspaper English* produced at the University of Uppsala by Erik Smittberg (Smittberg 2014). When the study was undertaken, Smittberg had hand transcribed around 320,000 words of newspaper material from the nineteenth century. Joulain discovered that 160,616 of Smittberg’s data was in the BLNCNC, allowing Joulain to test the presence, nature and consequence of errors in a sample of the raw OCRed British Library data.

The findings of Joulain relevant to our study are that:

- i. “the spread and average quality of the OCR may be expected to vary considerably depending on the year and publication of the sample considered” (Joulain 2017: 94);
- ii. “OCR errors do have an impact on the wordcounts” (Joulain 2017: 97);
- iii. The issue of OCR error cannot be easily overcome by technological means – attempts to address the issue using a spelling variation correction program (VARD, Baron 2011) produced results that were “about the same quality as the uncorrected data” (Joulain 2017: 152), while an experiment to rescan the data using a state-of-the-art OCR system (Overproof, Evershed & Fitch 2014) produced results that included a higher proportion of “real word” errors (e.g.

hot was read as *not*). Real word errors are potentially the most damaging sort of error that such data can produce as they actively mislead users – seeing *lnol* in a word list with a frequency count of 1 can easily lead to the conclusion that an error has occurred in OCR. But seeing *not* with frequency results inflated by OCR errors rendering *hot* as *not* is altogether more difficult to detect (see Joulain 2017: 162–165 for further details).

While these results clearly suggest caution in any automated approach to the use of a dataset like the BLNCNC, there are possible remedies available that would make such a dataset usable. Firstly, close reading of results clearly recommends itself as being essential in supporting the interpretation of such data. Yet, close reading on its own cannot be the only response to the problem – the scale of the data would defeat most analysts if they had to resort to close reading as their principal form of analysis. Usefully, the second remedy that we can apply comes from further insights Joulain applies regarding where and when close reading should be employed. The OCR errors impact most on infrequent terms. In fact, the impact of OCR errors on infrequent terms is catastrophic (Joulain 2017: 121). Conversely, with the more frequent types in the corpus we should be able to find sufficient good examples that an analysis could proceed based on those examples. Of course, one should be cautious about what is lost – a key concern is whether there is evidence that there is systematic skew, with certain types of examples being lost. If so, then the positive evidence we see, i.e. the word forms successfully retrieved, is in some sense not representative of the whole because the data that is lost is in some way systematically different. Joulain suggests that this concern is ill-founded. Most errors produced by the OCR are relatively random – high frequency types are therefore reasonably reliable guides to the overall use of a word. Time may be a confounding factor – for example, variation in print and ink quality over time may lead to the degree of error for a word form varying over time. Nonetheless, throughout the errors themselves will be random, meaning that, while the recovery of real word forms may vary over time because of issues like print quality (discussed in more detail in the following section), the nature of the OCR error at any point in time will remain random. By using these insights, Joulain showed that by excluding low frequency word types from her analysis she could improve the quality of the results retrieved from her corpus when compared to the gold-standard. For example, a frequency threshold of 10 reduces the number of errors in the types extracted from her corpus to 7% as opposed to 56% when no frequency floor is used (Joulain 2017: 89). So, using the insights that Joulain gained about the nature of the OCR data allows us to exclude the most egregious OCR errors and reduce the scope of close reading that we have to engage in to allow the data to be used.

Most importantly for the purpose of this study, Joulain also explored the extent to which collocation statistics that were run on the uncorrected data were unreliable

because of the OCR errors in the BLNCNC. She began by using the frequency floor she proposed, and experimenting with the use of a significance test and effect sized approach to collocation. She discovered that, by combining both a significance test (log-likelihood) and effect size measure (mutual information) with her proposed frequency floor when calculating collocation using a span of 5, her results minimised false positives (to 7.02% of collocates) while detecting no false negatives.¹ What of the ranking of the collocates? In a test she found a Spearman's rank correlation of .82 when comparing collocates in the BLNCNC and the gold standard corpus using a span of 5 (Joulain 2017: 129). Joulain (2017: 133) concludes that when calculating collocates based on such data: "Best practice [...] involves using a frequency floor, avoiding large spans, and using MI as a ranking statistic, in combination with LL as a significance threshold." Of course, given Joulain's finding the analyst also has to be aware that false positives may be an issue – we use concordancing to verify that the examples of collocations that we discuss in this chapter are not false positives produced by some unusually consistent OCR error, for example. Accordingly, for our investigations we use a collocation span of 5, Joulain's recommended frequency of co-occurrence threshold of 10 and the LogRatio statistic² that combines a test of significance with an effect size measure. With an understanding of how we can approach collocation in relatively noisy data established, let us turn to consider the next methodological question – how to guide analysis in a very large dataset.

3. Scale

Our approach to scale links to the previous discussion – we use collocation, in the form of so-called Usage Fluctuation Analyses (McEnery, Brezina & Baker 2019) in order to guide where our investigation of shifts in the discussion of SLAVE may occur. UFA is a technique that visualizes shifts in patterns of collocation around a word, representing the results as a curve. The curve represents change in the pattern of collocation through time. In what follows we will look at two UFAs from the *Liverpool Mercury*, one for *slave* and another for *slaves*. At this point our aim is

-
1. Note that we report the findings for a span of 5 here as this is a typical span for collocation studies. However, if we had been applying techniques which typically use longer (typically whole document) window spans, such as topic modelling (see Brookes & McEnery (2019) for a critical introduction) or word vectors (see Desagulier (2019) for a relatively accessible overview) we would have faced substantial problems as larger spans produced very poor results in Joulain's experiments. See Joulain (2017: 129) who concludes, "Large spans are best avoided" as the statistics she evaluated "become increasingly unreliable in OCR data, when using larger spans" (Joulain 2017: i).
 2. Developed by Andrew Hardie (see <http://cass.lancs.ac.uk/log-ratio-an-informal-introduction>).

only to show the graphs and explain how they can be read. The necessary historical background (factors which may influence the shifting pattern of usage) is presented in Section 5. The linguistic contextualization of these graphs (exactly what the words co-occurring with the word in question are and how they vary over time) is undertaken in Section 6 when historical context and linguistic analysis are used together to explore the representation of slavery in the *Liverpool Mercury* and to reflect back on how we may interpret the UFAs. Our goal in this section is both to demonstrate how, as abstract representations of usage shift, the graphs may be read and to outline the broad trends they suggest.

The UFA of *slave* in Figure 1 shows how a shift in a pattern of usage may be represented as a curve. The y-axis shows the degree of agreement between pairwise comparisons of patterns of collocation through time. This value varies between 0 (totally

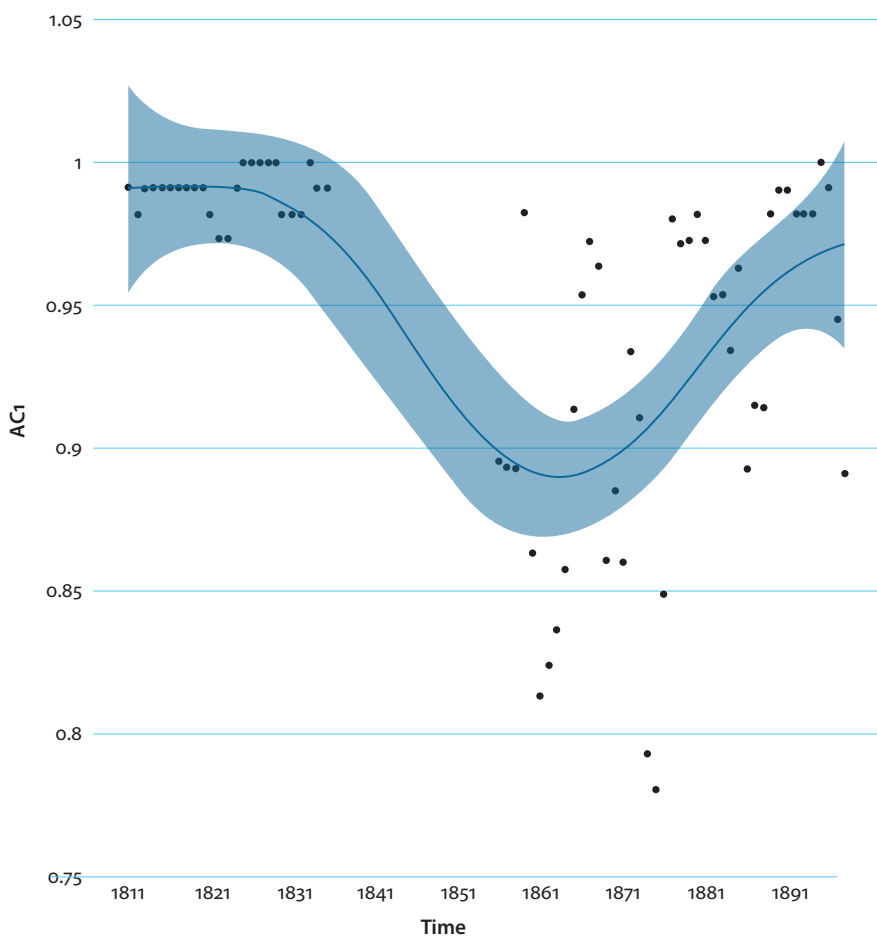


Figure 1. A UFA of *slave*

different) to 1 (identical). The curve shows the century beginning with a high degree of agreement in the pattern of collocation around *slave*, but then there is an accelerating divergence in the patterns of collocation around the word, represented by the downslope, starting in the late 1830s. This continues until the late 1850s when a trough is reached and the trend shifts to an accelerating pattern of convergence, with usage becoming increasingly similar for the rest of the century, as shown by the upslope.

A broadly similar pattern is observable in Figure 2, which shows a UFA of *slaves*. The century begins with the pattern of collocation around *slaves* being relatively stable, followed by accelerating divergence from the 1830s to 1850, followed by a period that is broadly (though in this case not exclusively) increasingly convergent thereafter.

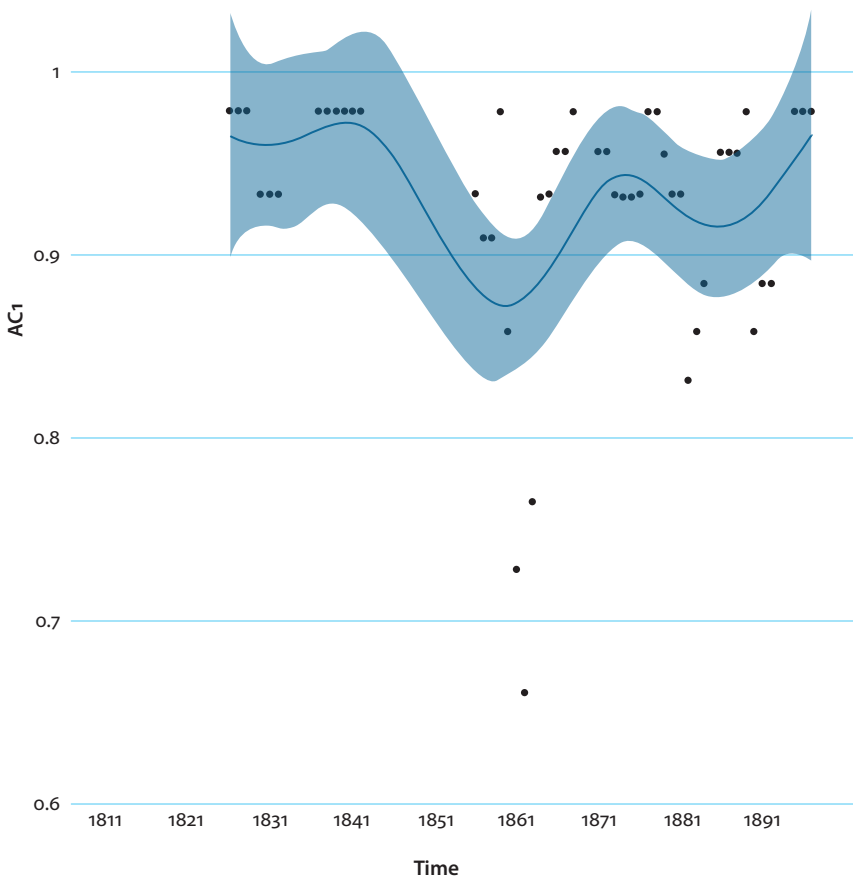


Figure 2. A UFA of *slaves*³

3. This UFA starts later than that in Figure 1, as there are insufficient examples of the word form *slaves* in the early decades of the *Liverpool Mercury* to produce a curve prior to 1825.

Readers interested in the mathematics behind an UFA are directed to McEnery, Brezina and Baker (2019). What directly concerns us in this section are two questions – how reliable are the calculations underlying the UFA, given the quality of the data we use, and how might we verify the reliability of the graphs?⁴ We must resolve these questions before trying to contextualise the UFA historically and explain the patterns of co-occurrence that produce it.

A UFA is based on a calculation of collocation for a given word – *slave* in Figure 1, *slaves* in Figure 2. So, the question of the reliability of these graphs is ineluctably bound to the question of the reliability of collocation calculations in data of this sort. Given that we have dealt with this issue in the previous section, and that the UFAs in Figures 1 and 2 were based on the collocation parameters we established in that section to minimise the issue of noisy data, we may be reasonably confident that these UFAs represent a fair ground truth of usage of these words in the data. However, UFAs may be more prone to error than a simple collocation because they pass through time. Joulain's results were not focused on how the variable nature of data through time may impact upon accuracy. So, for example, it may be the case that the divergence shown in the graphs in Figures 1 and 2 are caused not by change of usage per se, but by some deterioration in the quality of the OCR that produces increasing noise in the UFA. There are potentially multiple different reasons that OCR quality may vary, making this a difficult point to explore simply on the basis of the historiography of the newspaper – changes in paper quality, ink quality, formatting and fonts may all account for increases in OCR errors as may the more obvious issues of varying levels of damage to the originals (see Balk & Conteh 2011: 156–7; Holley 2009: 2–4; King 2005: 168; Tanner et al. 2009; and Alex & Burns 2014 for further discussion). It is beyond the resources of our investigation to thoroughly explore how these issues may impact our analysis – yet, studies such as those of Tanner et al. (2009) have shown that OCR quality varies markedly over time in any given title in the BLNCNC.

How, in such a context, might we verify that what we have seen in the graphs is a result of changes in usage rather than simply variations in OCR quality? The answer we explore in this chapter follows from the suggestions of Baker and Egbert (2016) and Egbert and Baker (2019) – we triangulate in our exploration of usage change. The UFAs here provide us with an initial impression of where change

4. A further issue may occur to readers – how do we deal with varying amounts of data through time? For example, the *Liverpool Mercury* produced 724,228 words of data in 1811, while it contains 47,043,434 words in 1900. UFA guards against the possibility that an increase in evidence may lead to an apparent shift in usage – see McEnery, Brezina and Baker (2019: 421–422).

occurs. We follow up on this using an earlier technique, from McEnery and Baker (2017) which uses collocations in blocks of n years over a long period of time to explore usage change. This result is much more qualitative than the UFA approach as the overall pattern is built up by successive investigations of sets of collocates through time. While the approach also uses automatically derived collocates, the categorization of those collocates into broad semantic fields, a mainstay of the approach, is conducted through the manual inspection of concordance lines in which the collocation occurs. This both allows the issue of OCR error to be investigated in a qualitative, yet focused, fashion while also providing another analysis of the same question that the UFA addressed, the usage of `SLAVE` over time, by a somewhat different method. Hence, through this triangulation we can explore the issue of OCR error and triangulate two analyses together.

With this established, we can now begin the process of the historical and linguistic interpretation of the UFAs presented. Let us begin that process by considering the historical background to this study.

4. Liverpool and the slave trade

The livings of most Liverpudlians were bound up with the slave trade, in some form or other, by the end of the eighteenth century (Howman 2007: 277). The city's links with the transport of enslaved Africans were acknowledged by contemporaries. In 1783, the brothers John and Thomas Hodgson, owners of slaving vessels, described the benefits that the slave trade had bestowed upon Liverpool: "the African trade has done unheard of wonders for this Town [...] particularly in these last 12 or 18 months, & it has a great deal of fortune yet in the wheel" (quoted in Morgan 2007: 15). A number of nineteenth-century historians, for instance, Richard Brooke (1853) and the anonymous writer, "Dicky Sam" (1884), chronicled the city's involvement.⁵ Morgan (2007: 18, 21, 33–34) has explained how Liverpool's investment in the slave trade was so effective due to the business acumen of its merchants – in exploiting new markets and maintaining good relations with African traders, its locational advantages, particularly in relation to its proximity to the Isle of Man, a tax-free haven prior to 1765, and by wider demographic and commercial factors, such as population growth and the creation of six wet docks.

The most successful slave traders from the city matched their economic success with social and political clout. Sherwood (2007: 40–44) has written about

5. Dicky Sam was a nickname used to describe people from Liverpool.

John Gladstone, father of the British Prime Minister William Ewart Gladstone. John Gladstone owned a number of sugar and coffee plantations in Demerara (now part of Guyana) and Jamaica but also served as a Tory member of parliament for three different constituencies between 1818 and 1827; partly owned the pro-Conservative *Liverpool Courier*; and published a number of pamphlets and tracts, some of which defended the trade in slaves. Gladstone justified the brutal suppression of a slave rebellion in Demerara, 1823, and, as a consequence, was gifted £1,400 and some expensive dinner service items from the city of Liverpool to acknowledge “his most important services to the town of Liverpool”.⁶

In the late eighteenth century, Liverpool MPs opposed the Slave Trade Act of 1788, introduced by William Dolben, which restricted the number of slaves that could be transported by a vessel in relation to its tonnage. The following year, sailmakers and bakers from the city spoke out against abolition (see Thomas 1998: 507, 514). Yet, a number of historians who specialise in Liverpool abolitionism have argued that opposition to the slave trade in Liverpool has been underestimated. Howman (2007) explains that abolitionist sentiment during the first campaign of the 1780s was muted, perhaps due to a fear of reprisals and an unwillingness to offend fellow Liverpoolians who were heavily involved in the slave trade. However, he references a number of brave individuals, such as Edward Rushton, the poet, bookseller and founder of the Liverpool School for the Blind, and the Anglican cleric, Henry Dannel, who openly espoused abolition from the 1780s. Howman argues that the work of these men enabled the city to take a central role during the campaign for emancipation, from the 1820s to 1834, when a number of groups, such as the Liverpool Society for Promoting the Abolition of Slavery and the Liverpool Ladies’ Anti-Slavery Association, actively opposed the trade. Baggett (2014) concurs that the work of the so-called Roscoe Circle in 1787–88, comprised of early abolitionists such as Rushton, the physician James Currie and the group’s namesake, William Roscoe, who became the city’s first pro-abolitionist MP in 1807, whilst initially not publicising its views openly, participated in the national movement by means of literary contributions. Baggett argues that the city turned its back on the trade from the mid-1790s and, by 1806, the majority of city residents no longer defended it. On 1 January 1876, the *Liverpool Post* declared: “there are matters on which they [the English people] are capable of being raised to a white heat of anger that burns up sooner or later the object of their hate. And Slavery is one of them” (quoted in Mulligan 2009: 190).

However, the work of Sherwood (2007) suggests that Liverpool’s involvement with the slave trade continued to flourish after abolition. She refers to the increase in

6. As quoted in Smithers (1825: 448).

trade in slave-grown sugar, tobacco and, particularly, cotton, which were unloaded in the city. Moreover, the leading abolitionist, Zachary Macaulay, believed that 36 vessels had sailed from Liverpool on slaving voyages between 1807 and 1809. Sherwood (ibid.) wonders to what extent customs officials turned a blind eye to such activity.

This historical context is the one in which our interpretation of the UFAs presented is framed. Further historical detail will be added in Section 6 as findings from the corpus demand.

5. The *Liverpool Mercury*

The British Newspaper Archive website describes the paper we base our study upon, the *Liverpool Mercury*, as a Liberal, reformist publication which circulated not only in Liverpool but in areas of Lancashire, Cheshire, Wales, the Isle of Man and London. First published in 1811, its coverage included wider domestic and international events but it specialised in reporting of local social issues.

The *Liverpool Mercury* was a large newspaper. After 1858, its publication changed from weekly to daily and a larger edition was published on Fridays. By the 1880s, it claimed that the weekly two pence edition of the paper contained “seventy-two long columns making it one of the largest newspapers in the world.”⁷

6. Analysis

In order to triangulate with our UFA analysis, we proceeded to analyse *slave* and *slaves* in the *Liverpool Mercury* with the method of McEnery and Baker (2017) – we explored the corpus in discrete ten year chunks, identifying collocates in each chunk. The space available to us here precludes a full analysis of SLAVE, so we will limit our study to a consideration of consistent collocates – collocates which appeared with SLAVE in every decade of the analysis.

There are four consistent collocates of SLAVE in the *Liverpool Mercury* corpora – *of*, *the*, *trade* and *who*. These occur in almost every chunk examined. In this study, we focus upon these consistent collocates, examining in what context they tended to co-occur with SLAVE and if this context was unchanging. They are important both for this study and the UFA as they represent a bedrock of similarity across the century.

We initially explored the instances in which *of* collocated with SLAVE using the early decades of the *Liverpool Mercury Corpus*. We wondered if the collocate

7. See <<https://www.britishnewspaperarchive.co.uk/titles/liverpool-mercury>>

of might often directly follow the term SLAVE and might indicate a possessive relationship, in constructions such as SLAVE of [an owner]. However, we found that *of* appears directly to the right of SLAVE in only 118 occurrences out of 2,505 in the *Liverpool Mercury* (in the period 1811–1840). In these instances, we found only two contemporary instances where a slave was identified in terms of belonging to an owner. For instance, an article of 3 March 1824 publicising an anti-slavery volume by the Liverpoolian author, John Taylor, referred to the “slaves of a planter” who transported his slaves from Tortola to Trinidad with disastrous results for the slaves concerned (*Liverpool Mercury*, 19/03/1824).

Much more common were references to where slaves were located. Places such as the West Indies, the United States, Trinidad, St. Domingo (Saint-Domingue) and Jamaica all featured. Significantly, an article of 26 February 1830, referred to “the white slaves of Manchester, Birmingham, or Leeds”. Indeed, Huzzey (2014: 365) has described how the position of the freed black slaves and British labourers were often compared during this period. We identified 17 further metaphorical references to slaves in the instances of SLAVE of – for instance, “slave of a guilty passion” (*Liverpool Mercury*, 03/11/1820), “slave of misery” (*Liverpool Mercury*, 26/10/1838), “slaves of popular opinion” (*Liverpool Mercury*, 24/06/1831) and “slaves of pride” (*Liverpool Mercury*, 06/03/1812). Thomas (1998: 453, 477) has discussed the tendency by eighteenth-century Britons and Americans to use slavery metaphors without reflection, believing that “they had little idea of the implications of what they were saying”.

The word *of* co-occurs with SLAVE throughout the century in the *Liverpool Mercury*, most often two places to the left of SLAVE. From 1811–1840, *of* took this position on 865 occasions out of 2,505, or 34.56% of the time. When we examined these 865 instances of ‘*of* + [word] + SLAVE’ we found that [word] was almost always *the* and, in 169 cases this pattern was followed by the word *trade*. The form of SLAVE appearing in the pattern is invariably *slave*. In other words, three of the four consistent collocates of SLAVE – *of*, *the* and *trade* – are also in a strong relationship with one another forming, in effect, a lexical bundle (Biber et al. 1999: 992–993), *of the slave trade*.

Note that, mindful of the potential of the OCR data to disrupt our analyses, we undertook a qualitative analysis of the results, reviewing each instance of the word which preceded the phrase *of the slave trade*. We explored these one by one to account for false positives and negatives. These decisions were made on a common-sense basis. So, for instance, the words *ebulition* and *abolitiow* were judged to be sufficiently close to *abolition* to be counted but the word *arbrditiow*, which may possibly have also been *abolition*, was not. This enabled us to consider more instances of the phrase *abolition of the slave trade* than were initially generated via concordancing, which only provided up instances which were unaffected by OCR errors. For instance, in the period 1811–1840 we counted 75 occurrences of

the phrase in the *Liverpool Mercury Corpus* (1811–1840) but concordancing only produced 56 instances. Nonetheless, overall, we did not find large scale disruption of patterns of collocation in any given period that we observed in terms of false positives, leading us to draw an initial conclusion that some of the fears we may have had about the validity of the UFAs were ill-founded.

An exploration of the instances of *of the slave trade* proved to be a useful introduction to the types of discussions the *Liverpool Mercury* was focussing on in reporting that referenced the slave trade. In almost all cases, this pattern is preceded by a noun. Hence, it is that pattern '[noun] + *of the slave trade*', that became the focus of our analysis.

With some faith in the UFA established, we decided to shift the blocks of analysis in our qualitative study to match three broad periods that seem common in the UFAs of *slave* and *slaves* – these words seem to produce three thirty year periods of activity: early (1811–1840), mid (1841–1870) and late (1871–1900). By aligning our investigation with those periods, we could gain a further advantage from this study, which was to gain insight into the nature of the change of usage across these broad periods. To do so we will look at the pattern '[noun] + *of the slave trade*'. Table 1 summarizes the occurrences of this pattern, sorted into semantic fields, which will be explored in detail in Tables 2–5. Table 1 gives a high level view of how instantiations of the pattern distribute across broad semantic fields in the corpus over time. The allocation of words to semantic fields was undertaken by a close reading of the examples by two analysts who then agreed on which semantic fields the nouns in question related to in the examples given using the same grounded approach to the assignment of words to broad categories of meaning used in McEnery and Baker (2017: 163–164).

Our exploration of Tables 2–5 is in part to seek broad patterns which may align with or challenge what the UFA suggests is happening. However, it is also a way of exploring a variability in the pattern itself: while *of* is a consistent collocate of *SLAVE*, is the use of the pattern '[noun] + *of the slave trade*', which drives much of the collocation of *of* with *SLAVE*, consistent itself? Might the apparent consistency of usage implied by the consistent collocation of the two be illusory, masking a more complex pattern of usage?

What is perhaps most striking about the findings of Tables 2–5 is that, generally, they reveal a consistency in reporting about the slave trade across the century. There are differences – in terms of the categories nouns are assigned to and how those nouns are divided between those categories – but they are mostly small. Let us begin by exploring the *CESSATION* semantic field.

6.1 The cessation of the slave trade

In the *CESSATION* category, Table 2 shows that discussions about the *suppression* or *abolition of the slave trade* dominate.

Table 1. An overview of the distribution of the semantic fields of nouns in the pattern '[noun] + *of the slave trade*' in the *Liverpool Mercury*

[noun] + <i>of the slave trade</i> in the <i>Liverpool Mercury</i>			
	Total examples Early Period (1811–1840)	Total examples Mid Period (1841–1870)	Total examples Late Period (1871–1900)
advocacy	5	13	8
cessation	93	255	212
condemnation	5	22	25
discussion	24	38	44
guilt	0	2	0
history	2	5	1
illegality	3	0	2
legislation	2	1	3
location	0	2	10
opposition	1	6	10
persistence	13	36	30

Table 2. CESSATION: instances of '[noun] + *of the slave trade*' in the *Liverpool Mercury* (1811–1900). Frequencies of occurrence greater than 1 shown in parentheses

[noun] + <i>of the slave trade</i> in the <i>Liverpool Mercury</i>			
	Early Period (1811–1840)	Mid Period (1841–1870)	Late Period (1871–1900)
CESSATION	abolition (75), suppression (14), destruction, prevention, rejection, renunciation	suppression (157), abolition (61), extinction (12), prevention (6), repression (4), cessation (2), destruction (2), extirpation (2), abandonment, annihilation, decline, discontinuance, downfall, extermination, prohibition, relinquishment, success ⁸	suppression (149), abolition (28), repression (12), extinction (6), prevention (5), checking (2), [death] blow, cessation, defeat, destruction, extirpation, mitigation, [break the] neck, overthrow, stoppage, withdrawal

8. Occurs in the phrase: "utterly despairing of the future success of the slave trade". See *Liverpool Mercury*, 18/07/1851.

In the early period of the *Liverpool Mercury Corpus*, relevant reports usually related to Britain's attempts to persuade other countries, particularly France, Portugal and Spain, to agree to abolition. This aligns with the work of Kaufmann and Pape (1999: 631, 634) who write that, for sixty years after the first abolition bill, Britain embarked on a crusade "to cajole, bribe, and, where possible, coerce the other slave-trading nations into compliance" in a policy they describe as the most expensive example of costly international moral action. The year 1814 has the largest number of these articles and they, alongside articles from 1815, almost always refer to the Congress of Vienna during which Britain pushed for a unified resolution on the issue. For instance, on 24 March 1815, a four-hour speech by Lord Castlereagh in the House of Commons was cited by the *Liverpool Mercury*: "He stated, that Spain and Portugal had agreed to the abolition of the Slave Trade, at the expiration of eight years; and expressed his hope, that, by negotiation, those Powers would consent to a still more contracted period".

Reports in the remainder of the early period, up until 1840, were less optimistic, with frequent references to resistance by Portugal and Spain to sign treaties (see, for example, *Liverpool Mercury*, 08/03/1839). The *Mercury* suggested that the British public felt impatient with the process of emancipation. On 28 March 1823, a letter was copied in its entirety from the *Derby Mercury* in which the writer lamented that the material condition of slaves had not improved since the abolition of the slave trade. The correspondent employed stark imagery and emotional language in order to get across their point:

We do not hear the groans of the slaves: we do not behold the uplifted arm of the insulting and brutalized driver: the smack of his whip does not resound in our ears... Our best feelings are not shocked by the sight of his lacerated body; nor are we horrified by exhibition of instances which have been attested to exist in the West Indies, of persons, in whose neglected wounds even maggots have bred. We do not witness the rupture of all ties of domestic relationship: we do not see the violation of all the tender charities of life; the child torn from its parents; and the wife carried away from her husband, to be subjected to the brutal lust of a tyrant.

Little more than a month later, on 2 May, a reference to public petitions by the inhabitants of Liverpool, Manchester and Newcastle for the abolition of the slave trade was briefly alluded to. This was one of the few references to popular mobilisation against the traffic of African slaves we came across in the *Mercury*.⁹

9. The campaign to abolish the slave trade has been called the first popular movement, characterised by mass meetings and petition campaigns. See Thomas (1998: 12) and Drescher (1994).

In the mid period of the corpus, the debate remained tethered to the subject of universal abolition but it widened in terms of the locations of countries under discussion. In reports concerning American abolition, the focus shifted from a report that the West Africa Squadron did not have the authority to search American ships¹⁰ to articles about bills introduced into the Washington House of Representatives “for the more effectual Suppression of the slave trade” (*Liverpool Mercury*, 01/05/1860). An increase in the slave trade in the Spanish colony of Cuba was reported on 27 July 1861 while Spain was condemned, the following year, for failing to honour her agreements: “A breach by Spain of the treaties into which she has entered with this country for the suppression of the slave trade is such a frequent occurrence as to attract very little attention” (see *Liverpool Mercury*, 26/06/1862). During this period, references to the establishment of agreements with African leaders increased, such as treaties with native chiefs in the Bight of Benin, reported on 13 April 1852. In 1858, two articles mentioned meetings which marked the fiftieth anniversary of the Slave Trade Abolition Bill (*Liverpool Mercury*, 31/05/1858 and 01/06/1858).

The geographical dimension of the debate shifted once again in the later period of the *Liverpool Mercury*, with drafts of treaties with Egypt and Turkey being reported (*Liverpool Mercury*, 10/03/1877). A number of articles referred to an alliance with the Sultan of Zanzibar, Barghash bin Said, who signed a treaty with Britain to abolish the slave trade in 1873 (see for instance, *Liverpool Mercury*, 21/06/1873 and 16/07/1875). It appears that the British determination to enlist African leaders in their campaign to suppress the traffic of slaves had political as well as moral motives. From the 1870s onwards, as Temperley (1972: 263–4) has written, European powers, including Britain, had extended their authority into Africa and they believed the practice of slavery was antithetical to the order they wished to impose on this new territory.

6.2 Discussion and location

In this section, we will discuss DISCUSSION and LOCATION. While these fields may seem separate initially, as will be shown, the two fields relate closely to CESSATION as the century proceeds, so we consider them together here.

In the DISCUSSION category, articles in the early period of the *Liverpool Mercury* revealed that the subject of the slave trade was frequently raised in both houses of Parliament. For instance, on 20 May 1825, a report stated:

10. *Liverpool Mercury*, 04/06/1858. The West Africa Squadron was established in 1808 in order to patrol the coast of West Africa – between 1820 and 1870, the squadron seized almost 1,600 ships and freed 150,000 slaves. See Walvin (2001: 265).

A number of papers have this session been laid before Parliament on the subject of the slave trade, and their recent publication, by order of the House of Commons, has excited a feeling of a most painful and disheartening nature in the minds of all that generous body of our countrymen, who have for years devoted their influence and exertions to the reduction of this most disgraceful and brutalizing traffic.

Table 3. DISCUSSION and LOCATION: instances of '[noun] + *of the slave trade*' in the *Liverpool Mercury* (1811–1900). Frequencies of occurrence greater than 1 shown in parentheses.

[noun] + <i>of the slave trade</i> in the <i>Liverpool Mercury</i>			
	Early Period (1811–1840)	Mid Period (1841–1870)	Late Period (1871–1900)
DISCUSSION	subject, state, effects, business, circumstance, discussion, enormities, question, specimen, talk	subject (13), state (4), question (2), enormities (2), existence (2), appearance, aspect, case description, difficulty, head, irregularities, parts, principle, position, purposes, room, talk, talking, whole	question (12), subject (11), case (3), existence (3), description/s (2), features (2), state (2), amount, condition, extent, meeting, portion, point, report, sections, view
LOCATION		place, stronghold	centre (2), stronghold/s (2), entrepôt, headquarters, hotbed, markets, routes, sources

The report went on to say that while England and America had done everything in their power to wipe out the stigma of slavery attached to them, France continued to allow traders to operate under its flag.

The *Mercury* also wrote about the meetings of abolitionist societies. France's involvement in the traffic of black Africans was raised in one such gathering, by the friends of the abolition of the slave trade on 1 July 1814. The article described Wilberforce as expressing dismay that the French colonies, taken by the British on the defeat of Napoleon, were to be restored to France: "The people in France were under an unhappy influence if they could suppose it advantageous to continue a traffic as contrary to common sense as it is to humanity..."

The *Liverpool Mercury* showed awareness, from the early nineteenth century, that slavery was not limited to black Africans. An article of 20 December 1820 stated:

Public attention has been several times directed of late to the traffic in Indian coolies, which reminds us of some of the worst features of the slave trade. A story now reaches us from the island of Réunion which effectively exposes the pitiful condition of the population annually shipped in cargoes from Hind[u]stan to the islands of the Indian Ocean.

The newspaper also printed several discussions about the West Africa Squadron, whose effectiveness had been in doubt since 1838, when a number of Quaker abolitionists had argued it should be withdrawn. Questions about the Squadron's capability intensified in the 1840s. A group of politicians led by William Hutt, MP for Gateshead, believed that the Squadron was too expensive and that, by antagonising trading partners such as France and the United States, it threatened Britain's economic wellbeing (see Thomas 1998: 657, 733). An article printed on 13 July 1858 cited Hutt's views that the Squadron had ultimately failed and should only be instructed to board and search vessels in cases of suspected piracy.

In the later period, the discussion shifted slightly. Alongside an article about the Slave Circulars (see *Liverpool Mercury*, 08/03/1876) – contentious instructions to captains that fugitive slaves should be returned to their former masters, providing they were in a country where slavery was legal – there was a muted reference to the Anti-Slavery Conference, which took place in Brussels in 1889–90, in an article about the Queen's Speech: "More self-congratulation will be expended over the meeting of the Slave Trade Conference at Brussels, and the hope is expressed that it will be found to be useful" (*Liverpool Mercury*, 11/02/1890).

As suggested in the discussion of the CESSATION field in Section 6.1, the shifting focus of CESSATION works to bring LOCATION and DISCUSSION together as geographical locations discussed in relation to the abolition of the slave trade altered as the century progressed. This phenomenon is verified by an exploration of articles in both the DISCUSSION and LOCATION fields. We found that the debate moved away from Spain and Portugal and their colonies in Cuba and Brazil, respectively, and countries such as Egypt and Turkey were cast into the spotlight. Several reports in the early period of the corpus and, to an even greater extent, in the mid period detailed attempts to persuade Portugal and Spain to sign anti-slavery treaties (see, for instance, *Liverpool Mercury*, 02/12/1814, 06/02/1818, 19/06/1818 and 19/07/1844). Again, a salient issue was the continuation of the slave trade in Cuba. For instance, on 23 September 1853:

The commercial advices from Cuba state that the question of the slave trade continues to give constant trouble to the official representatives of the British government in that island. Open violations of treaty are almost of weekly occurrence, and, under the active connivance of the Spanish authorities, the traffic is obviously increasing.

However, towards the end of the mid period, there were indications of achievement in terms of reports that the trade had ended in Brazil (see *Liverpool Mercury*, 29/06/1864). These reports were evidently premature because Brazil did not officially cease its trade in African slaves until 1888. These findings align well when we look more closer at the geographical aspect of reporting by examining articles within the LOCATION category. They also referred to locations, such as Zanzibar – "the

headquarters of the slave trade” and “the centre of the slave trade” (*Liverpool Mercury*, 11/01/1876 and 09/03/1885)¹¹ – and Lagos, the former “hotbed of the slave trade,” whose governor, at the end of the century, asserted it was now a safe and socially progressive society (*Liverpool Mercury*, 12/08/1897).

However, two articles in the LOCATION field, of 15 October 1852 and 12 October 1875, cited Liverpool as being a former stronghold of the slave trade. The second reference appeared in a letter by “Somerset’s Ghost” which condemned the city: “While other towns are protesting against the atrocities and illegal actions of the Tory Government, how is it that Liverpool is silent?” This is another reference to the Fugitive Slave Circulars, described by Mulligan (2009: 183) as “one of the most prominent issues in political debate in 1875 and 1876”. The Admiralty’s instructions to captains in 1875 that they should return fugitive slaves to their former masters provoked a slew of petitions in England, proving that the issue of slavery was still capable of eliciting a strong emotional response.

6.3 Persistence of the slave trade

The PERSISTENCE category leads us to articles which discuss the continuance of the slave trade. Reporters often adopted a tone of dismay: for instance, “Sierra Leone papers to the 24th June have been received, and they supply the most heart-rending accounts of the rigorous continuance of the slave trade” (*Liverpool Mercury*, 13/10/1826).

Table 4. PERSISTENCE: instances of ‘[noun] + of the slave trade’ in the *Liverpool Mercury* (1811–1900). Frequencies of occurrence greater than 1 shown in parentheses

[noun] + of the slave trade in the <i>Liverpool Mercury</i>			
	Early Period (1811–1840)	Mid Period (1841–1870)	Late Period (1871–1900)
PERSISTENCE	continuance (7), increase, progress, remains, remnant, restoration, revival	revival (15), renewal (7), continuance (2), extension (2), re-opening (2), continuation, infraction, modification, perpetuation, restoration, resumption, resuscitation, vestige	revival (7), increase (4), development (3), re-establishment (3), activity (2), continuance (2), spread (2), activity, perpetuation, power, resumption, return, vigour, violation

11. Temperley (1972: 264) has explained how Zanzibar was a “slave empire in the classic sense” and was still a politically significant entity towards the end of the nineteenth century.

However, some of the nouns in the PERSISTENCE category, such as *increase*, *extension* and *spread*, not only suggest the slave trade continued but that it was growing in strength in the nineteenth century. For instance, an article of 17 May 1822 referred to a meeting of an organisation that had been founded by British abolitionists in 1807 to create a safe home for freed slaves in Sierra Leone:

The sixteenth anniversary meeting of the African Institution was held on Friday last, at the Freemason's Tavern, when upwards of 1500 persons of the greatest respectability assembled in the Great Hall; the Duke of Gloucester in the chair. The report stated, that there had been, since the last meeting, an alarming increase in the slave trade, on the eastern shores of Africa, particularly in the river Bonny, and at Calabar: 190 slave-ships having entered the former, and 162 the latter, engaged in that odious traffic.

Words belonging to the PERSISTENCE category lead us to two current affairs issues that were discussed in the mid-century – the dispute over sugar duties and the American Civil War. The debate over sugar spanned back to the Molasses Act of 1733 which established a tax on imports of sugar and molasses from non-British colonies as a result of lobbying by plantation owners from the British West Indies. Huzzey (2010) has shown how disagreement about the sugar tax emerged in the early nineteenth century but he argues that both sides supported anti-slavery. Those who favoured equal taxes for sugar argued that state protectionist policies discouraged sugar producers from reform and that it was wrong to protect sugar but not other slave-grown produce. In stark contrast to the present day, sugar was presented as a necessity of life which working class people should not struggle to afford and it was believed to have health benefits in terms of “weaning the working classes away from alcohol” (Huzzey 2010: 364).

Many of these arguments were voiced in the *Liverpool Mercury* in an article of 7 April, 1843, which reproduced a document and petition of the Liverpool Anti-Monopoly Association verbatim. The Association denied that removing the high protective tariffs on imported sugar and molasses would encourage “the continuance of slavery, and contribute to the extension of the slave trade”. It firmly cast as the villains colonial planters who had received £20 million of public money for losses “which have never been sustained”, while deliberately deceiving the British government about the costs of slave labour. Abolitionists were presented as well-meaning but biased – they did not oppose slave-grown cotton or tobacco. The Association insisted that the poor British people, who were paying nearly double the market value for sugar and molasses, “such wholesome and nutritious articles of food”, were the real losers.

Another issue which was raised in the middle of the century involved the American Civil War (1861–65). In the letters page of the issue of 6 November 1862, a correspondent wrote in condemnation of the Southern States:

Some weak-minded people have imagined that the separation of the States would lead to the abolition of slavery. It is a well-known fact, and declared by all Southern leaders, that extension of slavery and the restoration of the slave trade in all its “pristine glory” are the sole objects of the Black Confederacy.

The contributor to the letters page had a point: it is almost certain that, had the Confederates won the war, slavery would have continued in the South and the trade routes in the North would have been re-established (see Thomas 1998: 766).

Articles about the persistence of the slave trade in the later period of the *Liverpool Mercury* mostly concerned the revolt in the Sudan of 1884 led by Muhammad Ahmad. General George Gordon, the former Governor-General, was sent to Khartoum in order to secure the evacuation of loyal personnel and soldiers but, without the approval of the British government, he proceeded to defend the city for almost a year. Gordon had been a very popular figure in Britain and this was in no small way connected with his determination to rid Sudan of slavery. An article of 13 February 1884, for instance, after the destruction of the garrison at Sinkat, cited the views of Earl Cairns, the Conservative statesman, who argued in the House of Lords that “before they decided upon the abandonment of the Soudan [the British government] ought to have taken some effectual precautions against the resumption of the slave trade there”.

6.4 Condemnation, opposition and support of the slave trade

The *Liverpool Mercury* employed emotionally charged language in writing about the slave trade which is reflected in the way in which the number of concordances in the CONDEMNATION category increase as the nineteenth century progressed. In the early period, articles described shocking incidents where slaves had been deliberately thrown from slave ships by crew members who wanted to avoid being detained with slaves on board. For instance, on 6 April 1838:

The Duke of Wellington, on Tuesday, in reference to the horrors of the slave trade, declared that he had been authorized to state, and that he did state to the congress at Vienna, which he had attended, that slaves were confined in casks with their irons upon them; that a number of casks had been floating past vessels which were in pursuit of the slave ships.

Table 5. CONDEMNATION, OPPOSITION and ADVOCACY: instances of '[noun] + of the slave trade' in the *Liverpool Mercury* (1811–1900). Frequencies of occurrence greater than 1 shown in parentheses

[noun] + of the <i>slave trade</i> in the <i>Liverpool Mercury</i>			
	Early Period (1811–1840)	Mid Period (1841–1870)	Late Period (1871–1900)
CONDEMNATION	horror/s (3), evils, miseries	horrors (15), evils (3), atrocities, decadence, inhumanity, injustice	horror/s (15), evil/s (6), abomination, implements ¹² , iniquities, victims
ADVOCACY	advocate/s (3), support, partisans	encouragement (3), defence (2), revivers (2), agents, favour, interests, protector, soul ¹³ , support	encouragement (2), defence, favour, interests, means, promotion, supporters
OPPOSITION	opposer	detestation (2), abhorrence, enemies, opponents, oppose	director (3), heroes (2), condemnatory, enemy, figures, superintendent, work
ILLEGALITY	prosecution, seizure, trial		licitness, prosecution
HISTORY	memoir, time	time/s (2), days, history, relic	history
LEGISLATION	reading (2)	proceedings	reading (2), ratifications ¹⁴
GUILT		reproach, responsibility	

The punishment of these persistent slavers was also documented. One article referred to Robert Thorpe, first chief justice in Sierra Leone, who presided over the trial of the Dutch slave trader, Samuel Samo in April 1812 (*Liverpool Mercury*, 10/03/1815).

By the mid and later periods of the *Liverpool Mercury*, the phrases *horror/s of the slave trade* and *evil/s of the slave trade* had become stock phrases, employed by reporters in order to demonstrate, very quickly and conveniently, their repulsion at the notion of the enslavement of humans. These reporters did not go on to elucidate why the trade was horrible; there was simply no need – by that time,

12. Refers to wooden yokes and iron shackles that were used as implements of the slave trade. See *Liverpool Mercury*, 02/08/1884.

13. The article in question refers to the Portuguese as being the “life and soul of the slave trade”. See *Liverpool Mercury*, 12/06/1851.

14. Occurs in the sentence: “The ratifications of the Slave Trade Convention between Turkey and Great Britain have been exchanged.” See *Liverpool Mercury*, 19/04/1880.

the British public overwhelmingly despised the slave trade and was well aware of the horrors it entailed. In the later period, three articles which used these phrases referred to works of art or literature which referenced the slave trade. For instance, a short report of 8 November 1890 mentioned an African sketch entitled ‘The Tale of a Tusk of Ivory’, in which, it is reported “some of the horrors of the slave trade are laid bare”. Another article in this later period, printed on 2 August 1884, referred to an Anti-Slavery Society meeting in which busts of the early abolitionists, Graville Sharp and Thomas Clarkson, who died in 1813 and 1846 respectively, alongside implements of the slave trade, such as timber yokes and iron shackles, were placed on display. The slave trade was passing into memory during this period – its early opponents were being honoured, it was being represented in art and literature, and objects used to enslave black Africans were regarded as museum artefacts.

However, particularly in the early period of the corpus, the *Liverpool Mercury* acknowledged that the slave trade was still supported by men with political power. A number of articles belonging to the ADVOCACY category referred to citizens of France having a particular interest in the continuation of the trade. An article of 29 December 1820 quoted a letter from a Bordeaux resident who lamented that the slave trade was being protected by some members of the French government: “I wish I saw anything in the state of our political affairs which afforded a better prospect for the future; but circumstances lead me to fear that the warmest partisans of the slave trade are about to come into power.”

France’s attitude towards colonial slavery was complicated due to the immense political upheaval the country experienced in this period. Slavery was abolished in 1794 under the First Republic, was re-established by Napoleon Bonaparte in 1802 and abolished once again by the Second Republic in 1848. This was followed by a period in which the French government “repurchased” former slaves and forced them to work in their colonies as indentured labourers for a period of ten years, essentially ensuring the continuation of slavery under another name. On 25 June 1858, an editorial in the *Liverpool Mercury* criticised this scheme. A second reference to this scheme appeared in 4 February 1859 in a speech by Queen Victoria which was printed verbatim:

I have great satisfaction in announcing to you that the Emperor of the French has abolished a system of negro emigration from the east coast of Africa against which, as unavoidably tending, however guarded, to the encouragement of the slave trade, my Government has never ceased to address to his Imperial Majesty its most earnest but friendly representations.¹⁵

15. Fears that the scheme by the French government to use indentured labour for their colonies in the West Indies would inevitably lead to a revival of the slave trade was discussed in

Other references to supporters of the trade in the mid-century were more diverse, ranging from citizens of England, European empires and kingdoms of Africa. On 26 February 1847, the paper published a one sentence reference to the Rev. Henry Palmer's pamphlet in defence of the slave trade which had just been published. The Kingdom of Dahomey was judged to be the "mainstay and support of the slave trade in the Bight of Benin" and the Portuguese were termed "the life and soul of the slave trade both in the Interior and on the coast" (13 May and 12 June 1861, respectively). Even more interesting was an article of 4 June 1862, entitled 'Liverpool Slave Traders', in which the newspaper acknowledged that the secret agents of the slave trade in New York, no longer able to operate in their own country due to a greater determination from the government in Washington to enforce its own laws, had established themselves in Liverpool. The article referred to the slave ship, *Nightingale*, which was fitted out in the city and asserted that "the port of Liverpool will not long be found a convenient starting place for slave ships when once the people of Liverpool have been made generally aware that it is so used". Indeed, the reporter rammed the point home that the people of Liverpool "are interested in discovering and bringing to justice ruffians who attempt to turn a British port into a slave-trading station". More naïve readers might assume that the city had had no previous association with the slave trade at all.

However, it would be unfair to accuse the *Mercury* of editing the city's history. An article of 5 December 1892, detailing W.E. Gladstone's presentation of the freedom of the city, printed his speech in full in which he made a clear reference to Liverpool's involvement in the traffic of Africans. This was a sensitive subject for the prime minister, of course, because it was well-known that his father was a major slave-owner. Gladstone's speech emphasised that English support of the trade had not entirely vanished:

[Liverpool] had risen to that eminence not wholly but in part through a traffic on which we all now look back with shame and sorrow, viz., through the participation in the slave trade. (Hear, hear). The only apology that can be made for Liverpool in that matter is that the opinion to which her trade gave practical effect was an opinion universal throughout the entire country... We may read now the appeals of parliamentary candidates to the voters beseeching them to return to Parliament supporters of the slave trade if they valued their families, their country, and even their religion.

In the OPPOSITION category, an article entitled 'The Bishop of Oxford on missionary work' of 3 November 1868, also discussed Liverpool's guilt. The Bishop

the House of Commons on 10 July 1857. See Hansard <<https://api.parliament.uk/historic-hansard/commons/1857/jul/10/threatened-revival-of-the-slave-trade>>. Also see Flory (2016).

declared that: “more than any other part of the world, perhaps, except one, Liverpool was connected with [the slave trade], and therefore, Liverpool should be connected with the act of undoing it.” He went on to say that Liverpoolians should not “waste their energy expressing their detestation of slave trade” but must give practical help in their support of the work of the missionaries in Africa.¹⁶

The size of the *OPPOSITION* category increased in the mid-century and then again in the later period of the *Mercury*. There were references to a book about the early heroes of the slave trade, such as Wilberforce, Buxton and Sharp (see *Liverpool Mercury*, 14/10/1878); a report about the civil servant, William Henry Wylde, the former superintendent of the Slave Trade Department, being awarded the Order of Saint Michael and Saint George (*Liverpool Mercury*, 03/04/1880); and of General Gordon, the “great enemy of the slave trade” (*Liverpool Mercury*, 02/08/1884) and a reference to the late Prime Minister, Lord Palmerston, “not merely superintending but entering into the drudgery of every part of the work of the slave trade department”. As we can see, some of these references are historical and they add to the notion that the slave trade was beginning to be viewed as a phenomenon of the past.

7. Conclusion

The analysis presented in this chapter is a clear demonstration of the usefulness of the work of Joulain (2017). Using her work as a guide to deriving collocations from our corpus, we were able to use an established method of corpus assisted discourse analysis, collocation, in order to undertake a diachronic, historically contextualised, corpus assisted discourse analysis using nearly two billion words of newspaper data. We supplemented that approach with close reading of results and where, as noted, we identified false positives or negatives, we amended our analysis accordingly. Importantly, however, in doing so we lost no patterns to our semantic field analysis, though we did add some. Accordingly, we can be confident that the analysis presents a strong baseline for the study of *SLAVE* in the *Liverpool Mercury*. In the unlikely event that a perfect version of the corpus becomes available, we would suggest that what would be seen is nuance added to the results reported here rather than a falsification of them.

16. The concordances in the small *GUILT* category, only present in the mid-century, do not refer to British culpability. Brazil, for instance, is singled out as being “deeply steeped in the responsibility of the slave trade” in an article about free trade of 14 May 1841.

Similarly, by using triangulation, we have gained greater confidence in the UFAs we showed in Figures 1 and 2. By exploring patterns of collocation in a more qualitative way, using the decade by decade analysis of McEnery and Baker (2017) we explored the impact of OCR on relevant collocates in the corpus and did not find cause to believe that variable OCR quality in any way invalidated the findings of Joulain (2017) on which our work on collocation and UFA in this chapter was based. In addition, by then combining the insights from the UFA with a period based analysis, we reorganized the corpus into three chunks which broadly aligned with the major periods of stability and change in the UFA graphs. In those periods we pursued our qualitative approach further, conducting a detailed study of one pattern, '[noun]+ *of+ the+ slave*' and gained further corroboration of the UFA while also gaining insight into the driver of the pattern shown for SLAVE in Figures 1 and 2. The pattern of variation within the exploration of the pattern broadly matched what the UFA was suggesting – the mid period is distinct from the early and late periods. The mid period seems to be linked to an increased and widening discussion of the slave trade relative to the early period and late period. It is possible to argue that the decline of that discussion in the late period is linked to the success of efforts to eliminate the slave trade, especially as we see the locations discussed shifting and narrowing. That, we would argue, may be an overall explanation for the UFAs in the chapter. A further corroboration of the UFA approach arises from this study – we considered the possibility that the link between *of* and *slave*, which is in part included within the UFA and is central to the exploration of the patterning which links the two, may in fact be unstable, with the pattern denoting different meanings through time. This was not found to be the case – a consistent collocation was based on a consistent pattern of uses, at least for the main pattern analysed. So the overall finding of the triangulation approach taken here, is that the UFA and qualitative analysis were mutually reinforcing – the former is suggestive of a broader pattern of usage, the latter adds detail and nuance that is supportive of that patterning and adds explanatory power to it.

That explanatory power is increased when we look at the collocates underlying the UFA. Our focus on consistent collocates masks another feature of the UFA – both *slave* and *slaves* are only associated with consistent and relatively transient collocates. If we consider *slave* as an example, there are no words which begin to collocate with it in the early part of the century and remain with it to the end (initiating collocates). Similarly, there are no collocates which link to *slave* for a long time from the beginning of the century but which then cease to link to it much later in the century (terminating collocates). Rather, in addition to the few consistent collocates, there are only transient collocates which attach for between one and fifteen years to *slave*. What we see in the early period are three transient collocates, none of which survive through to the mid period. The mid period then

has a process of divergence driven by the accumulation of 33 transient collocates, eight of which do continue into the beginning of the late period when further flux is introduced by an additional 41 transient collocates. The pattern for *slaves* is similar (though with one exception that will be noted shortly) – it has three transient collocates in the early period which do not continue into the mid period, 16 transient collocates in the mid period, three of which continue into the late period and 13 new transient collocates added in the late period. Unlike *slave*, it does, however, have one collocate which straddles all three periods and which may be characterised as being on the cusp of initiating – *were*, which initiates in 1839 and is consistent thereafter. This links well to the historiography; following the 1833 Abolition Act, discussions of slavery in the newspaper swiftly shift to being strongly linked to the past. This is as clear a linguistic marker of the changing usage of *SLAVE* in the nineteenth century as it is possible to imagine. Overall, the matrix underlying the UFA is an indication not only of the usage of *SLAVE* being in flux in the century but also of a disconnect between the early and the mid periods being greater than between the mid and the late period, as we may have surmised from both the qualitative analysis and the UFAs.

In terms of what our study of the newspaper itself shows, the *Liverpool Mercury* did not shy away from discussing the slave trade. Within its pages, Liverpool's past involvement in the traffic of enslaved Africans was acknowledged and writers conveyed a determination that any further participation must not be tolerated. One report of 12 October 1875, about the Slave Circulars, berated Liverpool for not speaking out to a greater extent but, on the whole, the newspaper gave the impression that the vast majority of the inhabitants of the city were united against slavery. Graphic and emotive descriptions of the horrors entailed by the slave trade were more likely to appear earlier in the nineteenth century. By the middle of the century, the phrases *horror/s* or *evil/s of the slave trade* were employed to express a perfunctory aversion to the trade.

Nineteenth-century reports in the *Liverpool Mercury* revealed how Britain struggled to persuade other countries, particularly other empires, to agree to abolition. Articles detailed an array of diplomatic overtures – draft treaties, international assemblies, and unconcealed bribes – which were often unsuccessful. The effectiveness of Britain's other approach, namely the operations of the West Africa Squadron and the difficulties it faced, were also frequently discussed. The geographical focus of the *Mercury* shifted as the century progresses. Earlier reports detailed the tergiversation and resistance of France, Spain and Portugal. In the middle of the century, the narrative altered to include the activity of American slavers, suspected of operating in Liverpool, and attempts to prevent the practice of slavery in African territories. All of these articles provide a fascinating insight into Britain's unending patience and perseverance to achieve universal abolition – an aspiration which, to nineteenth-century Britons, must often have looked like a lost cause.

References

- Alex, Beatrix & Burns, John. 2014. Estimating and rating the quality of optically character recognised text. In *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage, Madrid, Spain*, Apostolos Antonacopoulos & Klaus U. Schulz (eds), 97–102. New York NY: Association for Computing Machinery.
<https://doi.org/10.1145/2595188.2595214>
- Anon. 1884. *Liverpool and the Slavery: An Historical Account of the Liverpool-African Slave Trade by a Genuine Dicky Sam*. Liverpool: A. Bowker and Son.
- Baggett, Florence. 2014. The Slaving Capital in the Era of Abolition: Liverpool's Silent Rejection of the Slave Trade, 1787–1807. Master's Thesis, Victoria University of Wellington.
- Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, Paul & Egbert, Jessie. 2016. *Triangulating Methodological Approaches in Corpus Linguistic Research*. New York NY: Routledge. <https://doi.org/10.4324/9781315724812>
- Balk, Hildelies & Conteh, Aly. 2011. IMPACT: Centre of competence in text digitisation. In *HIP '11: Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, 155–160. Beijing, China. <https://doi.org/10.1145/2037342.2037369>
- Baron, Alistair. 2011. Dealing with Spelling Variation in Early Modern English Texts. PhD dissertation, Lancaster University.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
- Brooke, Richard. 1853. *Liverpool as it was During the Last Quarter of the Eighteenth Century, 1775–1800*. Liverpool: J. Mawdsley and Son.
- Brookes, Gavin & McNery, Tony. 2019. The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies* 21(1): 3–21. <https://doi.org/10.1177/1461445618814032>
- Desagulier, Guillaume. 2019. Can word vectors help corpus linguistics? *Studia Neophilologica* 91(2): 219–240. <https://doi.org/10.1080/00393274.2019.1616220>
- Drescher, Seymour. 1994. Whose abolition? Popular pressure and the ending of the British slave trade. *Past and Present* 143: 136–166. <https://doi.org/10.1093/past/143.1.136>
- Egbert, Jessie & Baker, Paul. 2019. *Using Corpus Methods to Triangulate Linguistic Analysis*. New York NY: Routledge. <https://doi.org/10.4324/9781315112466>
- Evershed, John & Fitch, Kent. 2014. Correcting noisy OCR: context beats confusion. Paper presented at the Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage, 45–51, Madrid, Spain. <https://doi.org/10.1145/2595188.2595200>
- Flory, Céline. 2016. Between indenture and slavery? African indentured laborers in the French West Indies (1852–1862). *Francophone Africa: Critical Perspectives*, University of Portsmouth. <<http://francophone.port.ac.uk/?p=1223>> (16 January 2020).
- Hitchcock, Tim. 2013. Confronting the digital: Or how academic history writing lost the plot. *Cultural and Social History* 10(1): 9–23. <https://doi.org/10.2752/147800413X13515292098070>
- Holley, Rose. 2009. How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitisation programs. *D-Lib Magazine*, 15(3–4). <<http://www.dlib.org/dlib/march09/holley/03holley.html>> (31 March 2021).
- Howman, Brian. 2007. Abolitionism in Liverpool. In *Liverpool and Transatlantic Slavery*, David Richardson, Suzanne Schwarz & Anthony Tibbles (eds), 277–296. Liverpool: Liverpool University Press.
- Huzzey, Richard. 2010. Free trade, free labour, and slave sugar in Victorian Britain. *The Historical Journal* 53(2): 359–379. <https://doi.org/10.1017/S0018246X10000051>

- Huzzey, Richard. 2014. The slave trade and Victorian “humanity”. *Victorian Review* 40(1): 43–47. <https://doi.org/10.1353/vcr.2014.0025>
- Joulain, Amelia T. 2017. *Corpus Linguistics for History: The Methodology of Investigating Place-name Discourses*. PhD dissertation, Lancaster University.
- Kaufmann, Chaim D. & Pape, Robert A. 1999. Explaining costly international moral action: Britain’s sixty-year campaign against the Atlantic slave trade. *International Organisation* 53(4): 631–668. <https://doi.org/10.1162/002081899551020>
- King, Edmund. 2005. Digitisation of newspapers at the British Library. *The Serials Librarian* 49(1): 165–181. https://doi.org/10.1300/J123V49N01_07
- Loy, Wesley. 1997. 10 Rumford Place: Doing confederate business in Liverpool. *South Carolina Historical Society* 98(4): 349–374.
- McEnery, Tony & Baker, Helen. 2017. *Corpus Linguistics and 17th Century Prostitution*. London: Bloomsbury.
- McEnery, Tony, Baker, Helen & Dayrell, Carmen. 2019. Working at the interface of hydrol-ogy and corpus linguistics: Using corpora to identify unrecorded droughts in nineteenth-century Britain. In *Using Corpus Methods to Triangulate Linguistic Analysis*, Jessie Egbert & Paul Baker (eds), 52–84. New York NY: Routledge. <https://doi.org/10.4324/9781315112466-3>
- McEnery, Tony, Brezina, Vaclav and Baker, Helen. 2019. Usage fluctuation analysis: A new way of analysing shifts in historical discourse. *International Journal of Corpus Linguistics* 24(4): 413–444. <https://doi.org/10.1075/ijcl.18096.mce>
- Morgan, K. 2007. Liverpool’s dominance in the British slave trade, 1740–1807. In *Liverpool and Transatlantic Slavery*, David Richardson, Suzanne Schwarz & Anthony Tibbles (eds), 14–42. Liverpool: Liverpool University Press.
- Mulligan, William. 2009. The fugitive slave circulars, 1875–76. *Journal of Imperial and Commonwealth History* 37(2): 183–205. <https://doi.org/10.1080/03086530903010343>
- Partington, Alan. 2003. *The Linguistics of Political Argument: The Spin-Doctor and the Wolf Pack at the White House*. London: Routledge. <https://doi.org/10.4324/9780203218259>
- Partington, Alan, Duguid, Alison & Taylor, Charlotte. 2013. *Patterns and Meanings in Discourse Theory and Practice in Corpus-assisted Discourse Studies (CADS)* [Studies in Corpus Linguistics 55]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.55>
- Prescott, Andrew. 2014. I’d rather be a librarian: A response to Tim Hitchcock, confronting the digital. *Cultural and Social History* 11(3): 335–341. <https://doi.org/10.2752/147800414X13983595303192>
- Sherwood, Marika. 2007. *After Abolition: Britain and the Slave Trade since 1807*. London: I.B. Tauris. <https://doi.org/10.5040/9780755622245>
- Sherwood, Marika & Sherwood, Kim (2007). *Britain, the Slave Trade and Slavery, from 1562 to the 1880s*. Kent: Savannah Press.
- Smithers, Henry. 1825. *Liverpool, its Commerce, Statistics, and Institutions: with a History of the Cotton Trade*. Liverpool: Thomas Kaye.
- Smitherberg, Erik. 2014. Syntactic stability and change in nineteenth-century newspaper lan-guage. In *Late Modern English Syntax*, Marianne Hundt (ed.), 311–330. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139507226.023>
- Tanner, Simon, Muñoz, Trevor & Ros, Pich H. 2009. Measuring mass text digitization quality and Usefulness. *D-Lib Magazine* 15 (7–8): 1082–9873. <<http://www.dlib.org/dlib/july09/munoz/07munoz.html>> (31 March 2021). <https://doi.org/10.1045/july2009-munoz>
- Temperley, Howard. 1972. *British Antislavery, 1833–1870*. Columbia SC: University of South Carolina Press.

- Thomas, Hugh. 1998. *The Slave Trade: The History of the Atlantic Slave Trade 1440–1870*. London: Papermac.
- Walvin, James. 2001. *Black Ivory: Slavery in the British Empire*. Oxford: Blackwell.
- Wodak, Ruth. 2001. The discourse-historical approach. In *Methods of Critical Discourse Analysis*, Ruth Wodak & and Michael Meyer (eds), 63–94. London: Sage.

CHAPTER 3

Diachronic change in the ordering of kinship binomials

A contrastive perspective

Anna Čermáková

Charles University, Prague

One of the specific characteristics of binomials is the ordering of their elements and the degree of their reversibility. A diachronic perspective suggests it is particularly the kinship binomials that show a strong unfreezing trend away from the male-first ordering. This study explores, diachronically and from an English-Czech contrastive perspective, kinship binomials in children's literature. It confirms earlier findings of a gradual diachronic reversal of term ordering in kinship binomials that extends across languages. However, at the same time, binomial sequencing seems to be a complex interplay of linguistic, cognitive and real-world influences. The diachronic reversal of preference in the ordering is limited to particular binomials and may be linked to a more general change in the discourse, namely the shift towards greater informality.

*Come, father and mother,
And sister and brother,
Come, all of you, turn to
And help one another.*

(Anna Sewell, *Black Beauty*, 1877, Chapter 35)

1. Introduction

The above extract from the 19th century classic *Black Beauty* by Anna Sewell contains two kinship binomials: *father and mother* and *sister and brother*. While *father and mother* is the canonical form in the 19th century, *sister and brother* would, in fact, more likely occur as *brother and sister*. This chapter explores binomials like these that occur in children's books and aims to contribute to a discussion on principles that govern the ordering of the nouns within binomials. Binomials

are coordinated same word-class word pairs. One of the specific characteristics of binomials and a question of interest for linguists has been the degree of their reversibility, that is, whether the order of elements is fixed or reversible. A diachronic perspective on the ordering of binomials has been until recently limited to a few individual examples of reversal of preference. One of examples is the binomial *father and mother*, which is the preferred form in the nineteenth century and undergoes a reversal in the twentieth (Moon 1998). Mollin (2014) in her extensive study of binomials in English observes that some gender binomials have undergone a change in reversibility and points out that it is particularly the kinship binomials that show a strong unfreezing trend away from the male-first ordering (Mollin 2014: 160–161), which is often taken to be one of the key binomial ordering constraints. This trend is, however, rather isolated and has not been yet fully explained. Binomial sequencing seems to be a complex interplay of linguistic, cognitive and real-world influences.

Kinship terms belong to the very basic relationship terms that designate social categories (Lévi-Strauss 1949). Kinship terms denoting family relationships are particularly relevant in children's books due to their high frequency of occurrence. Children's books constitute not only one of the important sources for developing reading and other literacy skills (Jerrim & Moss 2019) but also present to young readers ideological norms of our society (see e.g. Reynolds 2011: 34). This study aims to explore kinship terms and specifically kinship binomials in this specific text-type and it aims to do so from an English-Czech contrastive perspective in order to establish whether the above-mentioned change of the order of the two nouns within a binomial is a specifically English phenomenon or whether it is extendable beyond the monolingual perspective. From the existing cross-linguistic studies of binomials it is clear that different languages may prefer opposite orders for otherwise comparable binomials (e.g. see Levin & Ström Herold in this volume for a contrastive study of English, Swedish and German, Ebeling & Ebeling (2013) for Norwegian and English, Giammarresi (2010) for English and Italian, or Pordány (1986) for Hungarian). Klégr and Čermák (2008) is an example of a contrastive study of Czech and English focusing on the translation of binomials in Shakespeare, Chaucer and Beowulf. None of these studies looks specifically at gendered or kinship binomials. This study, therefore, aims to fill in this particular gap and examine the reversibility of gender binomials cross-linguistically, focusing on binomials that have shown a strong unfreezing trend in English.

Section 2 presents a theoretical background, Section 3 presents the data and methodology used and discusses the degree of the data comparability between the two languages. Section 4 presents and discusses the analysis of the English data

and Section 5 of the Czech data. Section 6 extends the cross-linguistic discussion and Section 7 is the concluding section.

2. On binomials and kinship terms

The term ‘binomial’ is usually attributed to Malkiel (1959). Sandra Mollin in her comprehensive corpus-based study of English binomials defines binomials as “coordinated word pairs whose lexical elements share the same word class” (Mollin 2014: 1). Binomials are part of the language’s phraseological repertoire where some belong to idioms and may also serve as a basis for analogous word combinations (Čermák 2010). Mollin (2014: 30) also notes that binomial word classes are highly register-specific. In certain text-types, such as English legal texts, the use of binomials has a long tradition and binomials are very frequent (Gustafsson 1975, 1984; Hiltunen 1990), for example, *law and order*, *liable and responsible* or *null and void*. Stylistically, binomials are a special case of parallelism with the main functions of coordination and intensification and, as Danet (1985: 283) observes, the legal register is “striking for its use of elaborate parallel structures”.

In English, the most frequent are nominal binomials (Mollin 2014: 29). There are various classifications of binomials. Malkiel (1959) distinguishes five categories: (i) A and B are near-synonyms; (ii) A and B are mutually complementary; (iii) B is the opposite of A; and less frequently (iv) B is a subdivision of A (or vice versa); and (v) B functions as a consequence of A. However, as Gustafsson (1975) points out, the semantic categories have fuzzy boundaries. Linguistically, one of the most intriguing characteristics of binomials is their differing degree of reversibility. This has been conceptualised as a cline: from reversible to irreversible with some in between with preferential ordering. Most of the studies seem to agree that “irreversibility rather than reversibility is the norm” (Mollin 2014: 43, see also Benor & Levy 2006: 236; Gold 1991: 104; Gustafsson 1976: 631, but cf. Malkiel 1959, see also Lohmann 2014).

Another, related, area of research has focused on the various phonological and semantic constraints on the ordering of the binomial elements. Sobkowiak (1993: 395) postulates a general semanto-phonological principle for ordering in English: UNMARKED-BEFORE-MARKED. The ordering constraints may be metrical-phonological (e.g. Jespersen 1905) and semantic, e.g. the social status of referents, chronological sequence, positive before negative, desirable before undesirable, the more important before the less important, the light before the dark, masculine before feminine (Abrahams 1950) etc. Cooper and Ross (1975) postulate altogether twenty semantic principles. One of the most frequently mentioned

constraints is iconicity (Mollin 2014: 79), which “predicts that if the real-world referents of the two binomial elements are perceived to be in some chronological or cause-and-effect order, the elements will retain this order in the binomial.” In other words, as Croft (2003: 102) says, “the structure of language reflects in some way the structure of experience”, a point relevant also for this study.

Two semantic constraints particularly relevant for this study can both be subsumed under the already mentioned UNMARKED-BEFORE-MARKED principle: ‘male before female’ and ‘power source before power recipient’ (Mollin 2014: 65). In both of these cases the ‘male’ is considered to be the unmarked gender (Mollin 2014: 69). To what extent this holds across gendered binomials in children’s books will be explored in the following sections. The power constraint is more complicated than the gender constraint. According to Mollin (2014: 81), it predicts “that the linguistic element whose real-world referent is the more powerful or central in the society appears first”. It builds on speakers’ (in our case also readers’ and writers’) extra-linguistic knowledge of the social status hierarchies in their respective cultures, which may be a less straightforward matter in the case of an immature child reader (see Nikolajeva (2014) for the concept of ‘novice’ reader).

Benor and Levy (2006) propose a constraint of ‘perceptual markedness’ which includes the intuitive notion that one term of a concept pair is more “central” or “primary” than the other (see Mollin 2014: 80). The kinship binomial *Mum and Dad* thus instantiates a clash between several constraints, the gender and power constraints on one hand and perceptual markedness on the other. From the perspective of gender and power the male term, *Dad* in this case, should come first but from the perceptual markedness perspective, it is the female “because the mother is typically more central in the parental role” (Mollin 2014: 95). Similarly, Dant (2013) observes that family relationships constitute a semantic constraint that overrides the power constraint.

Recently, the ordering of binomial elements has also been repeatedly linked with their frequency of occurrence and thus their establishment in the mental lexicon. For example, Siyanova-Chanturia et al. (2011) report a significant correlation of eye-tracking measurements with the frequency of the binomial sequence. Morgan and Levy (2016) explore the idea whether the preferred order is driven linguistically (semantically, phonologically and lexically) or whether prior experience and exposure with the specific items may have an influence. Based on their experimental modelling, they reduce the list of the constraints to seven (in the order of their weightings): iconic sequencing (i.e. elements existing in sequence are also ordered in this sequence, e.g. time sequence), perceptual markedness, formal markedness, power, final stress (i.e. the final syllable of the second word should not be stressed), length (i.e. the shorter word comes first), and frequency (the more frequent word comes first).

However, as Mollin (2014: 116) points out: “[s]urprisingly, the question of the historical development of reversibility is virtually uncharted territory in studies on binomials” except perhaps as part of studies on legal register, see, for example, Hiltunen (1990).¹ Some individual cases of a diachronic change in the binomial order have been documented. For example, Moon (1998: 153–154) notices the above mentioned change in the ordering of the *mother and father* binomial. As Mollin (2014: 125 and Figures 5.9 and 5.10 on page 132) documents, the change in the ordering of *mother and father* is historically quite remarkable and “certainly shows the most extensive change in reversibility that any binomial may undergo”.

Mollin (2014: 160) hypothesises that the reasons for change in this and other related gender binomials “lie in sociocultural changes, so that a weakening of a power differences in society is reflected in falling (ir)reversibility scores” (see Section 3 for the calculation of ‘irreversibility score’). Mollin shows, in addition to unfreezing observed for *mother and father*, also unfreezing trends for previously frozen binomials such as *men and women* or *brothers and sisters* (see also Motschenbacher 2013: 226). Mollin (2014: 160–161) concludes that “gender binomials thus have witnessed an exceptionally dramatic history of reversibility changes” with particularly two groups of gender binomials emerging: kinship binomials (juxtaposition of two kinship terms differentiated by gender) and general reference binomials differentiated by gender. The unfreezing trend in the case of *mother and father* is “extremely strong” away from male-first ordering. However, as Mollin also points out, the trends she observes (see Mollin 2014: 161, Figure 5.29) clearly show that the beginning of the change long precedes the feminist movement to which, she initially hypothesised, the change would be linked.

Additionally, there is another research strand highly relevant for the current study that considers ‘accessibility’ as a specific predictor of ordering. Onishi, Murphy and Bock (2008) suggest that more prototypical members of categories tend to be produced before less-prototypical members. Morgan and Levy (2016) show that prior knowledge and experience of specific binomial expressions has an influence on the way they are expressed. Iliev and Smirnova (2014) link the ordering in binomials with the psychological and demographic characteristics of a speaker, such as gender, race, geographic location, or religion. They find converging evidence that word order is in these cases linked to a broad set of features associated with the speaker; so, for example, the overwhelming majority of their

1. For a comprehensive study of the historical development of binomials in English see Kopaczky and Sauer (eds) (2017).

test respondents produced ‘male’ before ‘female’ and out of those participants who produced the reversed sequence, ‘female’ before ‘male’, the majority were women. There is a number of other studies that have found that speaker sex influences the ordering (e.g. Hegarty et al. 2011; Wright & Hay 2002; Wright, Hay & Bent 2005). Tachihara and Goldberg (2019) in their cross-linguistic study of English and Japanese specifically focus on the ordering of proper names, which are often excluded from studies on binomials, and conclude that in both languages cognitive accessibility is the most important factor in the ordering of familiar name binomials, that is, speakers of both languages tend to produce the name of the person they feel closer to before the name of the other member of the couple. Though proper name binomials may not strictly, in the phraseological sense, be considered binomials, their production is highly relevant to the question of reversibility and particularly to kinship binomials considered here.

3. Data and method

For the purpose of this study, children’s literature is defined in a simplified way, externally, as literature written for or marketed at children (for an overview of children’s literature see e.g. Reynolds 2011). Two corpora of English children’s literature (see Table 1) are used in this study: a 19th century children’s literature corpus (ChiLit) and a corpus of contemporary children’s fiction (texts published after 2000) created from texts in the *Oxford Children’s Corpus* (OCC).² ChiLit corpus is a 4.4 million word corpus (Čermáková 2017, 2018). The OCC subcorpus (12.9 million words) is a sample of mainstream contemporary books published for children by Oxford University Press after the year 2000.

There are no directly comparable data sets for Czech. Therefore, the analysis draws on data from the large databank of Czech language, specifically SYN_v7³ corpus but also other diachronic resources available from the Czech National Corpus. The earliest fiction books in Czech intended for young audiences available from this corpus date back to the beginning of the 20th century. The trend for the 19th century Czech is therefore established with the SyD tool (see Section 5) which is based on texts not intended primarily for children. The core of the

2. The ChiLit corpus can be explored at <<http://clit.bham.ac.uk/>>. I would like to thank the Oxford University Press for allowing me to access their data. The OCC corpus is searchable through SketchEngine <<https://www.sketchengine.eu/>>.

3. <https://en.wikipedia.org/wiki/Czech_National_Corpus>. The data is freely available at <www.korpus.cz>.

analysis uses two data-sets: children's books published before 1980 (JUNpre1980) and after 1980 up to present (JUNpost1980). The dividing date, 1980, is arbitrary. These subcorpora were created from SYN_v7 from fiction texts aimed at child audiences. The first data set is fairly small, consisting of 1.1 million tokens,⁴ which represents 14 books by 13 authors. While most of these books are texts written by Czech authors, the data set also includes one translation (Tolkien's *Hobbit*). This subcorpus covers for the purpose of this study a transition period between the 19th century (for which we do not have data) and contemporary language. It aims to show whether there are any changes in the trends this study investigates.

The second data set is larger, consisting of 7.2 million tokens. 25% of these are books by Czech authors and the rest are translations (with eight source languages) with the biggest proportion (59%) of translations from English. This reflects the situation on the Czech children's fiction market. The subcorpus contains 168 books by 194 authors (with 62 translators). Admittedly, the corpus is not balanced as the aim was to have as big data as possible – there may be some author / translator bias as some authors and translators are overrepresented, for example, among the authors Enid Blyton accounts for 9% of the texts in the corpus and Karen McCombie for 6%. Among the translators, Petra Klůňová accounts for 6% of the translations. The issue of a possible translation effect will be briefly discussed in Section 6, where a micro analysis of translations will be presented. The data used for this analysis come from the parallel corpus *InterCorp*,⁵ the subcorpus used consists of 1.3 million tokens of children's literature in English and their translations into Czech.

Table 1. Main corpora used for the study

Corpus	Language	Words	Time period
ChiLit (19th century children's literature)	English	4.4 mio	1826–1911
OCC subcorpus (contemporary children's literature)	English	12.9 mio	after 2000
JUNpre1980	Czech	ca. 0.7 mio	1907–1979
JUNpost1980	Czech	ca. 5 mio	1982–2013

For this study, Mollin's (2014: 1) definition of binomials as “coordinated word pairs whose lexical elements share the same word class” (see Section 2) is adopted.

4. This figure includes punctuation, which is approximately 30%.

5. Available from <www.korpus.cz>, more info at <https://en.wikipedia.org/wiki/Czech_National_Corpus>.

Generally, binomials are considered to belong to the language's phraseological repertoire and as with any phraseology, it is sometimes difficult to draw a clear line between phraseological and free-word combinations. Frequency, therefore, is often a good indicator. Binomials as such may not be directly comparable across languages and various other linguistic constraints need to be considered, e.g. English is a language with a fairly fixed word order, while Czech, in comparison, belongs to the so called "free word order" languages. In Czech, binomials corresponding to two English nouns linked with the conjunction *and*, have two basic forms: they are linked with the conjunction *a* ('and') and, less frequently, with preposition *s* ('with'), for example, *otec a matka* ('father and mother') but *chleba s máslem* (*bread and butter*, literally 'bread with butter'). In this study, the focus is, in both languages, on nominal binomials coordinated with the conjunction *and* and its Czech counterpart *a*.

The kinship binomials in this study are further contextualised around the concept of 'people' nouns. These are nouns that refer to people (e.g. *mother*, *king*) or people-like characters (e.g. *fairy*). Proper names are excluded. Thus, binomials that include at least one 'people' noun and that are linked with the conjunction *and/a* are considered. Binomials have been generated from all four corpora: ChiLit, OCC, JUNpre1980 and JUNpost1980. From ChiLit, they have been manually selected from all the trigrams with the cut-off point of 5 occurrences (1.13 words per million); from the remaining corpora the binomials have been extracted with CQL queries. The cut-off points are arbitrary. For the OCC subcorpus, the cut-off point is 18 occurrences, which corresponds to the same normalised frequency as for ChiLit. For the Czech data, for JUNpre1980 the cut-off point is 2 occurrences as the corpus is small (corresponding to 1.82 pm) and for JUNpost1980 it is 4 occurrences (corresponding to 0.55 pm). Czech is a highly inflected language and the binomials have therefore been generated as lemmas, which will be indicated by the use of small caps. For further analysis, binomials including kinship terms and selected gendered binomials, i.e. binomials consisting of nouns of the opposite gender, have been considered.

To establish the trends of reversibility, the relevant binomials have been checked for their reversed order variants. For selected binomials, irreversibility scores have been calculated (Mollin 2014: 40) as follows:

$$\text{irreversibility score} = \frac{\text{freq.}}{\text{freq.} + \text{rev. freq.}} \times 100$$

Thus, for example, *father and mother* occurs in the 19th century data 172 times and *mother and father* occurs 7 times. The irreversibility score is 96% ($172 / (172 + 7) * 100$), which can be also read the other way round – as the reversibility score, which is 4%.

4. Diachronic development of English kinship terms

4.1 Kinship binomials in the 19th century children's literature

There are 53 'people' binomials in ChiLit with a frequency of five (1.13 pm) or more occurrences. The most frequent one is *father and mother* occurring 172 times followed by *Mr and Mrs* (105 times), and *king and queen* (84 times). The majority of these binomials are of the *father and mother* type with two referents, three out of these 53 are of an appositional type, that is, both nouns refer to one person, for example, *son and heir*. Most of these binomials consist of nouns in the singular form, 16 are in the plural form, the most frequent one being *brothers and sisters* with 54 occurrences. Four consist of nouns of non-agreeing number, for example, *wife and children* or *brother and sisters*. In terms of gender ordering, there are 25 binomials that are not gendered, that is, they do not consist of nouns with both genders involved, for example, the most frequent one being *women and children* (39 occurrences). Of the gendered binomials, only four have the female noun as the first noun in the pair – *ladies and gentlemen*, *bride and bridegroom*, *mother and father* and *girls and boys* – all, except *ladies and gentlemen* (occurring 35 times), occur with very low frequencies: eight, seven and five times respectively. 25 out of the 53 'people' binomials are kinship binomials. The kinship terms included in this study cover "blood" and "family" relations and include as follows (in alphabetical order): *aunt*, *brother(s)*, *children*, *daughters*, *father(s)*, *mamma*, *mother(s)*, *papa*, *sister(s)*, *sons*, *uncle*, *wife/wives*. 11 out of these 25 are binomials that consist of nouns that include both genders, see Table 2.

Table 2. Kinship binomials in the 19th century children's literature corpus (ChiLit corpus)

Binomial	Freq.	Gender of the 1st noun	Freq. of the 1st noun	Freq. of the 2nd noun	Freq. of the reversed order
father and mother	172	M	3251	3193	7
brothers and sisters	54	M	386	418	3
brother and sister	31	M	1196	683	3
uncle and aunt	38	M	944	595	6
brother and sisters	17	M	1196	418	0
papa and mamma	15	M	1073	624	1
father and mamma	13	M	3251	624	2
fathers and mothers	12	M	65	89	2
sons and daughters	10	M	205	95	0
mother and father	7	F	3193	3251	172
aunt and uncle	6	F	595	944	38

Table 2 shows the frequency of occurrence of these binomials. Most of the binomials are in the singular, three in the plural and one where one of the nouns is in the singular and the second in the plural (*brother and sisters*). Most of the binomials refer to core family members: four of them refer to parents, three to siblings, and one to children. Two binomials, in fact the same binomial with both orderings, refer to more distant relations (*uncle and aunt*). Table 2 also shows in the last column whether the binomial occurs in the reversed variant. Highlighted in grey background are two binomial pairs which occur in both orders with a frequency above the set threshold, albeit in both cases the male-first variant is clearly dominant. There are further five binomials (*brother and sister*, *brothers and sisters*, *papa and mamma*, *father and mamma*, *fathers and mothers*) where the occurrence of the reversed variant was attested. Additionally, the table includes another variable that is thought of possibly affecting the ordering of the nouns within the binomials: the frequency of occurrence of the individual nouns. Most of the binomials follow the pattern of the first noun also being more frequent generally, with four exceptions: *brothers and sisters*, *fathers and mothers*, *mother and father* and *aunt and uncle*. The first two binomials follow the order male noun first. Both of them are also attested in the reversed order – with, however, extremely low frequencies: three and two occurrences respectively. The binomial *mother and father*, occurring seven times is actually the reversed order binomial of the canonical form *father and mother* occurring 172 times and *aunt and uncle*, occurring six times, is the reversed order binomial of *uncle and aunt* occurring 38 times.

4.2 Kinship terms in contemporary children's books

In the OCC subcorpus, in contrast to ChiLit, the most frequent binomial is *mum and dad* with 1002 occurrences.⁶ There are only 20 'people' binomials in OCC with a frequency of 18 or more occurrences (which corresponds to the normalised frequency of 1.13 pm). All of these are coordinated binomials of the *father and mother* type. There are eleven binomials in singular, eight in plural form and one where one of the nouns is in the singular and the second in the plural – *mother and brothers* with 19 occurrences. In terms of gender ordering, there are five non-gendered binomials, the most frequent one, *women and children* (with 57 occurrences), is the same as in the ChiLit corpus. Seven binomials have the female noun as the first noun in the pair – they are all kinship binomials with the exception of

6. Some of the kinship terms, such as *mum* or *dad* often occur capitalised. The overall frequency of occurrence of *mum and dad* is 1002, the number of non-capitalised occurrences is only 242. Therefore, the following counts will include both capitalised and non-capitalised occurrences.

ladies and gentlemen occurring 61 times. Eight binomials have the male first ordering with *men and women* and *king and queen* (occurring 130 and 152 times respectively) being the most frequent ones. Ten out of the 20 ‘people’ binomials belong semantically to kinship terms. The kinship nouns included in these binomials are (in alphabetical order): *aunt*, *brother(s)*, *dad*, *grandad*, *grandma*, *granny*, *grandpa*, *father*, *mum*, *mother*, *sister(s)*, *son*, *uncle*. In comparison with ChiLit, instead of *mamma* and *papa* we find *mum* and *dad* and in addition we also find *granny and grandpa* and *grandma and grandad*. Nine out of the ten kinship binomials consist of nouns that include both genders, see Table 2.

Table 3. Kinship binomials in contemporary children’s books (OCC subcorpus)

Binomial	Freq.	Gender of the 1st noun	Freq. of the 1st noun	Freq. of the 2nd noun	Freq. of the reversed order
mum and dad	1002	F	10250	9528	20
mother and father	143	F	7291	8865	43
brothers and sisters	76	M	1013	721	13
granny and grandpa	57	F	1161	827	0
brother and sister	47	M	2804	2207	1
father and mother	43	M	8865	7291	143
aunt and uncle	26	F	2415	3030	16
grandma and grandad	23	F	163	740	0
mother and brothers	19	F	7291	1013	0

Table 3 shows the frequency of occurrence of these binomials. Most of the binomials are in singular form. The last column shows whether the binomial occurs in the reversed variant. Highlighted in grey background is the binomial pair which occurs in both orders with frequency above our set threshold – *mother and father*, though in this case, unlike in ChiLit, the female-first variant is the dominant. All the binomials were attested in the reversed order, except *granny and grandpa*, *grandma and grandad* and *mother and brothers*. The last one is semantically asymmetrical and thus perhaps less relevant. There was only one occurrence of *sister and brother*, which suggests that female-first ordering in this binomial is very unusual. The table also includes the frequencies of the individual nouns that make up the binomials. Here again, most of the binomials follow the pattern of the first noun also being more frequent generally. Three exceptions to this rule are the binomials *mother and father*, *grandma and grandad* and *aunt and uncle*. Based on this data, binomials consisting of the nouns *mother*, *father*, *mum*, *dad*, *aunt*, *uncle* and informal terms for grandparents prefer the female-first ordering while *brother*

and sister retain the male-first constraint. The binomials *mother and father* and *aunt and uncle* show an unfreezing trend already in ChiLit (see Table 2).

The irreversibility scores were calculated for the gendered kinship binomials that occur in both data sets, see Figure 1.

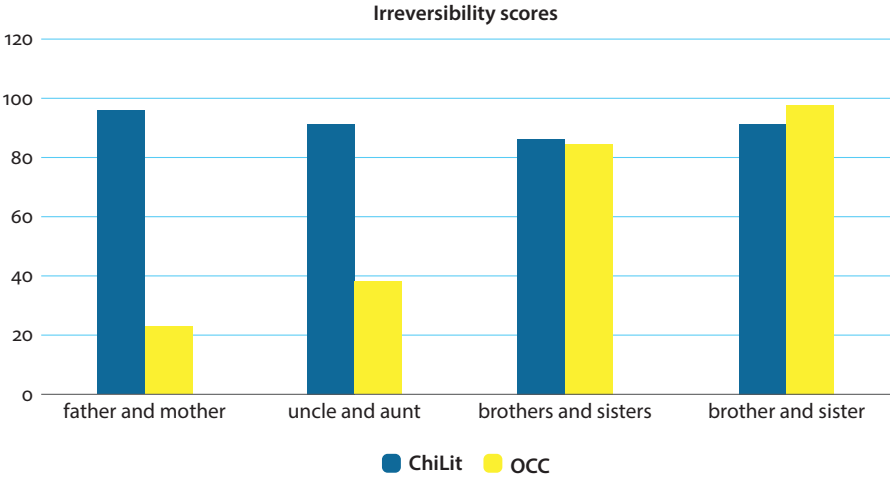


Figure 1. Comparison of irreversibility scores for gendered binomials in the 19th century (ChiLit) and contemporary texts (OCC)

The binomials in Figure 1 are in the male-first order, their canonical form in the 19th century. *Father and mother* and *uncle and aunt* show a strong unfreezing trend; in fact the scores show that it is the opposite version that is common in the contemporary data. The scores for *brother and sister* and *brothers and sisters* are very high and similar across both datasets, which suggest this ordering is nearly frozen. The score for the singular form *brother and sister* is even higher in the contemporary data. The diachronic comparison (not included in the figure) also shows that there are binomials that have virtually disappeared (*papa and mamma*) but also newly emerged binomials with the strikingly frequent *mum and dad*. *Mum and Dad* has an irreversibility score of 98, making it nearly frozen. The binomial that has changed the ordering most dramatically is, in line with Mollin's (2014) research, *mother and father* both in the singular and the plural. Figure 2 visualises this change in the OCC data.

Another noticeable change across the data is the decline in formality: *mum and dad* emerges as the most frequent 'people' binomial and only the informal variants for referring to grandparents occur in the OCC as binomials (*grandma and grandad* and *granny and grandpa*). There is also a decline in the frequency of binomials such as *wife and children*, *son and heir* and *husband and father* which can be perceived as rather formal.

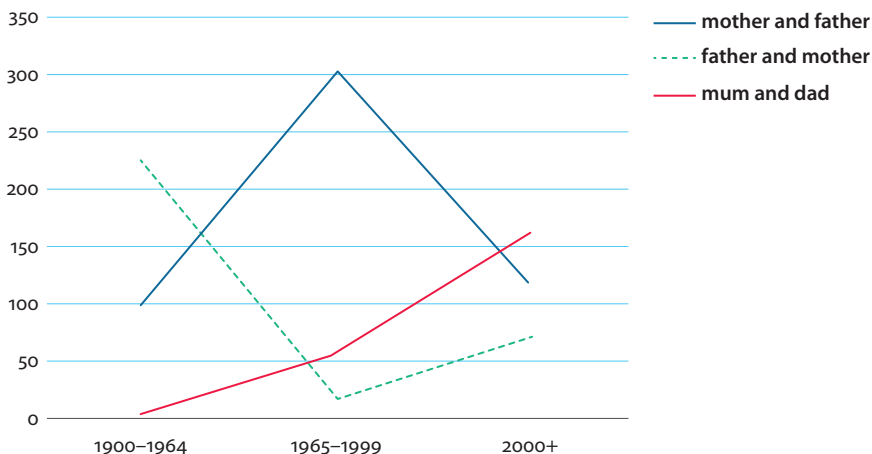


Figure 2. Change in the ordering of binomials referring to parents (OCC data)

5. Diachronic development of Czech kinship terms

Czech children's literature in today's sense started developing only in the 20th century (see e.g. Chaloupka & Voráček 1979). As there is no suitable data available for the 19th century, to establish an approximation for the norm of ordering for *otec a matka* ('father and mother') in 19th century Czech, I use the SyD tool.⁷ Figure 3 visualises the diachronic development of this binomial in 19th century Czech. The norm that this figure shows is male-first ordering and it is thus comparable to the English *father and mother*.

Czech is rich and very productive in the area of diminutives. Czech diminutives are derived morphologically and are very frequent in spoken register but also less formal text types. They are also typical in the language aimed at children. Many of the lemmas that will be discussed below are in the diminutive form. For the core binomial of this analysis, 'mother and father' / 'Mum and Dad', there are several Czech equivalents: *matka a otec* which is the most formal variant and corresponds to *mother and father*; *maminka a tatínek* which is a diminutive form but still rather formal; *máma a táta* which is an informal frequent variant corresponding to *Mum and Dad* and *mamka a tatka* which is a diminutive form and least formal of all the variants (attested only in the JUNpost1980 data set).

In JUNpre1980, the most frequent binomial is *TATÍNEK A MAMINKA* ('daddy and mummy') (34 occurrences). There are 50 'people' binomials in JUNpre1980

7. <<https://syd.korpus.cz/>>

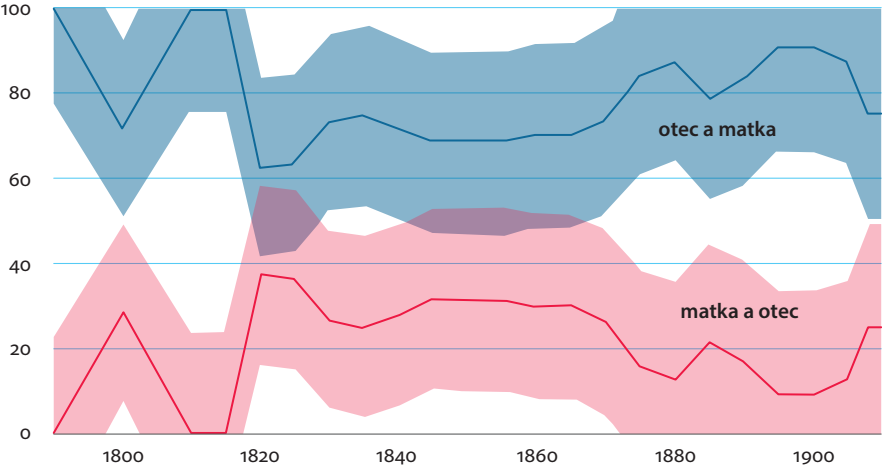


Figure 3. Visualisation of ‘father and mother’ (top line) and ‘mother and father’ (bottom line) in 19th century Czech

with the frequency of two (1.82 pm) or more occurrences. The majority of these are coordinated binomials of the ‘tatínek a maminka’ type. The lemma counts include both singular and plural forms but the distribution of the individual word forms (cases) shows that singular forms are more frequent (exceptions will be noted). See Table 4 for an example of word form distribution for the binomial TATÍNEK A MAMINKA.

Table 4. Distribution of the individual cases (word forms) for the TATÍNEK A MAMINKA binomial (JUNpre1980)

	Word from	Freq.		Word form.	Freq.
<i>tatínek a maminka</i>	sg. nom.	10	<i>tatínkové a maminky</i>	pl. nom.	7
<i>tatínka a maminku</i>	sg. accus.	5	<i>tatínky a maminky</i>	pl. accus.	1
<i>tatínkovi a mamince</i>	sg. dat.	5	<i>tatínků a mamínek</i>	pl. gen.	1
<i>tatínkem a maminkou</i>	sg. instrum.	4			
<i>tatínku a maminko</i>	sg. vocat.	1			

In terms of gender ordering, there are 16 gendered binomials, i.e. they consist of nouns of both genders. Of these, five have the female noun as the first noun in the pair – MAMINKA A TATÍNEK (‘mummy and daddy’, 14 occurrences), HOLKA A KLUK (‘girl and boy’, 4 occurrences), SESTRA A BRATR (‘sister and brother’, 2 occurrences), DÁMA A PÁN (‘lady and gentleman’, 2 occurrences) and MÁMA A TÁTA

(‘Mum and Dad’, 2 occurrences). The first three binomials are less frequent than their reversed variants, i.e. *TATÍNEK A MAMINKA* (‘daddy and mummy’) with 34 occurrences, *KLUK A HOLKA* (‘boy and girl’) with 16 occurrences and *BRATR A SESTRA* (‘brother and sister’) with seven occurrences are the more common variants. The binomial *DÁMA A PÁN* occurs in the plural and corresponds to ‘ladies and gentlemen’, which, similarly to English, seems to be one of the exceptions violating the male-first constraint.⁸

There are only eight kinship binomials, six of these are gendered (see Table 5). The kinship terms included in these binomials are (in alphabetical order): *bratr* (‘brother’), *máma* (‘Mum’), *maminka* (‘Mum’), *neteř* (‘niece’), *sestra* (‘sister’), *synovec* (‘nephew’), *táta* (‘Dad’), *tatínek* (‘Dad’). In comparison with the English data, these nouns do not include terms for ‘aunts’ and ‘uncles’ but this relationship is present through ‘niece’ and ‘nephew’.

Table 5. Gendered kinship binomials in JUNpre1980

Binomial	Freq.	Freq. of the reversed order	Gender of the 1st noun	Freq. of the 1st noun	Freq. of the 2nd noun
TATÍNEK A MAMINKA (‘daddy and mummy’)	34	14	M	1136	2876
MAMINKA A TATÍNEK (‘mummy and daddy’)	14	34	F	2876	1136
BRATR A SESTRA (‘brother and sister’)	7	2	M	105	144
SYNOVEC A NETERŘ (‘nephew and niece’)	3	0	M	5	11
SESTRA A BRATR (‘sister and brother’)	2	2	F	144	105
MÁMA A TÁTA (‘mum and dad’)	2	0	F	191	457

Table 5 shows the frequency of occurrence of these binomials. The second column shows the frequency of the reversed variant. Highlighted in grey are binomial pairs which occur in both orders. The table also includes information on the frequency of occurrence of the individual nouns on their own. The frequency constraint, i.e. the more frequent noun comes first in the pair, based on this small data set, seems not to be applicable.

8. As one of the reviewers has pointed out, this binomial may actually obey the power constraint, which is more obvious in the trinomial *lords, ladies and gentlemen*.

In JUNpost1980, there are 87 ‘people’ binomials with a frequency of four (0.55 pm) or more occurrences; the most frequent one being ‘man and woman’ followed by ‘woman and child’. In terms of gender ordering, there are 51 gendered binomials. Of these, 22 have the female noun as the first noun in the pair. With the following exceptions, most of them belong to kinship terms. The exceptions include the binomial ‘ladies and gentlemen’, ‘girl and boy’ (however, occurring more frequently in the reverse order), the reversed ‘woman and man’, the text-type specific ‘fairy and elf’ (VÍLA A SKŘÍTEK⁹ occurring equally frequently in the reversed order), ‘friend/fem. and friend/masc.’ (KAMARÁD A KAMARÁDKA occurring in the plural form with the male first ordering as more frequent) and ‘woman/wife and son’ (ŽENA A SYN).

There are 35 kinship binomials, of these 26 are gendered; 14 have the female-first ordering and 12 male-first. Of these 26, nine are pairs of reversed variants. In Table 6 only the more frequent variants are included. As the table shows there are five binomials referring to parents, four referring to grandparents and two referring to ‘aunt and uncle’ differing in various degrees of formality. In addition to ‘brother and sister’ there is also a binomial referring to cousins.

Again, the frequency constraint, that is, the more frequent element of the pair occurs as the first, is not convincing. Based on Table 6, the kinship binomials are more frequently ordered with the female element as the first one. All these female-first binomials involve parents, grandparents and ‘aunts and uncles’. This is similar to English as shown in Table 3 in Section 4.2. The kinship binomials with male-first ordering include in addition to ‘brother and sister’ and ‘son and daughter’, which are again similar to English, also ‘cousin/masc. and cousin/fem.’ and the most formal variants of ‘father and mother’ (OTEC A MATKA) and ‘uncle and aunt’ (STRÝC A TETA). However, frequency differences in these two cases are not significant, showing a strong unfreezing trend. The last binomial with male-first ordering is semantically asymmetrical (‘dad and grandma’) and thus perhaps less relevant for this comparison.

Figure 4 visualises the irreversibility scores of gendered kinship binomials occurring more than 7 times. Reading the figure from left, the first five bars show scores for binomials referring to ‘mother and father’. The first four variants, diminutives and less formal variants, have the female-first ordering. The last variant, the most formal OTEC A MATKA, is more frequent in male-first ordering but is also showing the strongest unfreezing trend. The following two bars refer to

9. Here the gender opposition is based on morphology, i.e. *víla* is feminine and *skřítek* is masculine – this, however, can be different in textual realisation, particularly where translations are concerned.

Table 6. Gendered kinship binomials in JUNpost1980

Binomial	Freq.	Freq. of the reversed order	Gender of the 1st noun	Freq. of the 1st noun	Freq. of the 2nd noun
MAMINKA A TATÍNEK (‘mummy and daddy’)	41	17	F	3259	1346
BABIČKA A DĚDA (‘grandma and grandpa’)	27	15	F	2104	605
MÁMA A TÁTA (‘mum and dad’)	26	18	F	2200	2328
BABIČKA A DĚDEČEK (‘grandma and grandpa’)	17	9	F	2104	793
MAMINKA A TAŤKA (‘mummy and daddy’)	13	3	F	3259	841
MAMINKA A TÁTA (‘mummy and dad’)	11	1	F	3259	2328
TETA A STREJDA (‘aunt and uncle’)	7	1	F	1054	474
BÁBINKA A DĚDEČEK (‘grandma and grandpa’)	6	4	F	102	793
BÁBI A DĚDA (‘grandma and grandpa’)	5	3	F	97	605
BRATR A SESTRA (‘brother and sister’)	47	12	M	2184	1538
OTEC A MATKA (‘father and mother’)	19	17	M	3390	3003
BRATRANEC A SESTŘENICE (‘cousin/masc. and cousin/fem.’)	15	5	M	202	214
STRÝC A TETA (‘uncle and aunt’)	7	5	M	468	1054
TAŤKA A BABIČKA (‘dad and grandma’)	4	0	M	841	2104
SYN A DCERA (‘son and daughter’)	4	0	M	1751	1154

grandparents and, while having the female-first ordering, also show a degree of reversibility. The next two bars refer to ‘aunt and uncle’. The first of these, TETA A STREJDA, is the less formal variant with female-first ordering and is more frozen than the formal STRÝC A TETA with a strong unfreezing trend similar to OTEC A MATKA. The last three binomials have the male-first ordering: ‘son and daughter’, ‘brother and sister’ and ‘cousin/masc. and cousin/fem.’

While it is difficult, due to the lack of data, to establish the nature of the diachronic development, the results suggest that such a development has in fact

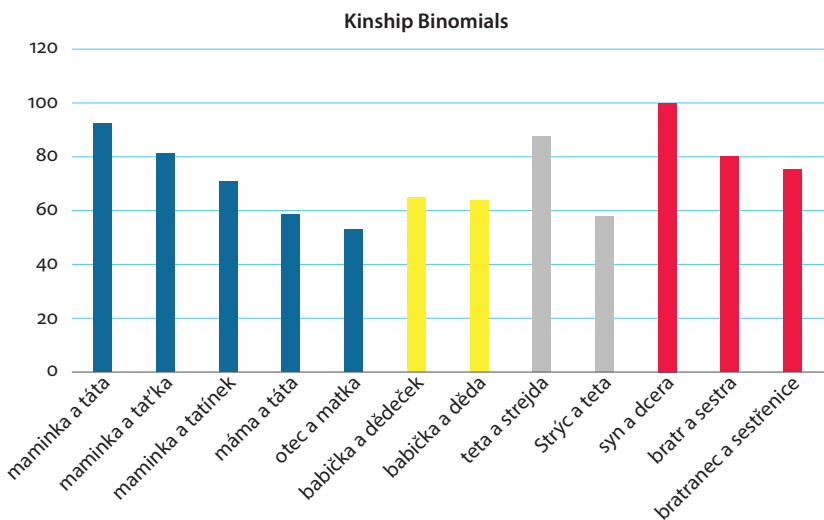


Figure 4. Comparison of irreversibility scores for gendered kinship binomials (JUNpost1980)

occurred. JUNpre1980, the transitional data set, suggests that there is an unfreezing trend for the ‘mother and father’ binomial, while JUNpost1980 confirms this trend. These results clearly point in two directions. Firstly, the unfreezing trend is the strongest for the binomials referring to parents, grandparents and ‘aunts and uncles.’ Secondly, the degree of formality seems to influence the ordering, i.e. the most formal pairs, OTEC A MATKA and STRÝC A TETA, tend to occur with male-first ordering more frequently.

I have analysed three of the variants of the binomial ‘mother and father’ in a subcorpus of the biggest available corpus of Czech SYN_v7¹⁰ to investigate whether a similar trend can be observed in other text-types than children’s literature, see Table 7.

Table 7. The distribution of three Czech variants for ‘mother and father’ across SYN_v7 corpus (frequencies normalised per million words)

Female-first ordering	Freq.	Male-first ordering	Freq.
MAMINKA A TATÍNEK	0.33	TATÍNEK A MAMINKA	0.19
MÁMA A TÁTA	0.22	TÁTA A MÁMA	0.11
MATKA A OTEC	0.39	OTEC A MATKA	0.49

10. The subcorpus consists of 272 mil. tokens of fiction, drama, autobiographies and popular non-fiction.

Table 7 confirms that the informal and diminutive variants prefer the female-first order (these differences are statistically significant at $p < .01$) while the formal variant OTEC A MATKA is preferred in male-first order (this difference, however, is not statistically significant).¹¹

6. Extended cross-linguistic perspective

In this section, I will look in detail at the binomials that have in both languages, from a diachronic perspective, changed the order of their elements, that is, ‘father and mother’, or shown a strong unfreezing trend, that is, ‘uncle and aunt’. While in the 19th century the irreversibility score for *father and mother* (male-first ordering) is nearly 100, indicating thus this form is frozen, in the post 2000 texts (OCC) the order is reversed with an irreversibility score nearing 80, which shows the form is fixed towards the female-first order with, however, some flexibility. *Mum and Dad*, which does not occur in the 19th century texts, has also an irreversibility score nearing 100, so again indicating this form is nearly frozen. The case is different for *aunt and uncle*. The 19th century strongly prefers male-first ordering (with an irreversibility score well over 80), while the contemporary texts prefer female-first ordering with a fairly low irreversibility score of 60 suggesting a greater degree of flexibility for the noun ordering.

For Czech, the contemporary texts show that the most fixed form is the female-first informal TETA A STŘEJDA (admittedly the frequencies of occurrence are low here), followed by diminutive MAMINKA A TATÍNEK with an irreversibility score of 70, followed by female-first informal variant MÁMA A TÁTA with an irreversibility score of 60. For the formal OTEC A MATKA, the male-first order is the preferred one, showing, however, a great degree of flexibility, having the irreversibility score 52. Also, the formal STRÝC A TETA (‘uncle and aunt’) prefers the male-first order, although again the frequencies of occurrence are too low for this binomial to make any conclusive statements. See Table 8 for an overview. One of the reasons for the generally lower irreversibility scores in Czech may be the linguistic typology of the two languages. Czech has a more flexible “free” word order and, as previous research shows, formulaic language is perhaps less pervasive: “[the idiom principle] operates in Czech to a different extent and allows for much greater variability” (Čermáková & Chlumská 2017: 93).

Another factor that may possibly be influencing the word order is the fact that the Czech JUN corpora contain a substantial proportion of translations from

11. <<http://sigil.collocations.de/wizard.html>>

Table 8. Irreversibility scores across time and languages

	ChiLit		OCC		JUNpost1980
father and mother	96	mum and dad	98	MAMINKA A TATÍNEK (‘mummy and daddy’)	71
uncle and aunt	86	mother and father	77	MÁMA A TÁTA (‘mum and dad’)	60
papa and mamma	94	aunt and uncle	62	OTEC A MATKA (‘father and mother’)	52
				TETA A STREJDA (‘aunt and uncle’)	88
				STRÝC A TETA (‘uncle and aunt’)	58

English (see Section 3). A very brief probe to see whether this may be the case was carried out on a small parallel corpus of children’s literature (1.3 million instances),¹² in which English source texts are aligned to their Czech translations. There are 12 occurrences of *mother and father* and they have all been translated as *matka a otec*, that is, the order of elements is the same. There are four occurrences of *father and mother*, which have been three times translated as *otec a matka* (keeping the order of the elements) and once the binomial was omitted altogether. The case of *mum and dad*, which occurs 23 times, is more varied. Most of the translation solutions (19) keep the female-first ordering using various variants – *máma a táta* and *máma s tátou* are the most frequent ones – two solutions reverse the order, there is one substitution with the word *rodiče* (‘parents’) and one omission. This suggests that the ordering in Czech for these may well be influenced by the source language.

Mollin’s (2014: 161) suggested explanation for the dramatic reversal in kinship terms being linked to social changes and the changing position of women is not entirely satisfactory, not least because, as she admits, the observed trend long precedes the beginnings of feminist movement. The “mother-first” ordering is also explained by the conceptual markedness constraint (see Section 2). Arguably, mothers and mums are more central and prototypical in children’s lives. The *aunt and uncle* binomial may be then constructed analogically. The same principle does not work for the other kinship binomials: *brothers and sisters*, *sons and daughters*, where the male-first ordering is preserved and the binomials are thus governed by different constraints. I would like to propose another possible reason for this difference. Based on the analysed data, the gradual diachronic reversal of preference in gender ordering of the kinship terms, and specifically the binomial

12. <<https://wiki.korpus.cz/doku.php/en:cnk:intercorp>>

father and mother, may be also linked to a more general change in children's literature, namely the shift towards greater informality, for example, while in the 19th century the most frequent family terms would have been *father* and *mother* (both as a binomial and individually), the by far most frequent binomial in contemporary texts is *Mum and Dad*. This is also supported by the Czech data where female-first is preferred with the more informal variants.

The reversal towards the female-first ordering is most visible in the binomials semantically consisting of lexemes denoting parents, grandparents and aunts and uncles. Unlike *brother and sister* where the male-first ordering is the norm, the words *Mum, Dad, mother, father* are often used as if they were proper names. Consider, for example, the frequent capitalisation of these nouns, e.g. *mum* occurs in the OCC 10,250 times and is capitalised in 8,221 (80%) cases. In comparison, *sister*, is capitalised in only 6% of its occurrences. *Mum* occurs only 27 times with determiners (14 times with *the* and 13 times with *a*)¹³ and the word is not very frequently premodified with adjectives – there are only 13 adjectives that occur with *Mum* more than twice. These make up 36 occurrences altogether (the most frequent one being *old* occurring 6 times).¹⁴ Again, the picture is slightly different for *sister*. *Sister* occurs with determiners 90 times (67 times with *a* and 23 times with *the*).¹⁵ *Sister* is also much more frequently premodified with adjectives – there are 36 different adjectives occurring twice or more which make up 316 occurrences. Grammatically, the lexeme *Mum* (and similarly *Dad*) behaves as a blend between proper name and common noun, while *sister* or *brother* less frequently do so. They are, in this sense, inherently different; see some vocative uses in examples (1-2) from the spoken component of the BNC, which show their proper name like behaviour. If *sister* or *son* were to be used in similar context, it would be stylistically marked.¹⁶

- (1) **Mum, mum** watch this, are you watching ...
- (2) Nana's delighted **Mum!** isn't she? **Mum! Mum!** She'll have to read these books I said **Mum!** nothing about dinosaurs. Yes, darling?

In children's stories, mums and dads are defined by being somebody's mums and dads, often we do not even learn their names (unlike with siblings). They are *Mum*,

13. In addition, *Mum* occurs in 19% of occurrences with a possessive pronoun or possessive 's.

14. Only L1 position is considered in this search.

15. *Sister* occurs with a possessive pronoun or possessive 's in 64% of occurrences.

16. I would like to thank one of the reviewers for pointing out that *son* may be used in similar contexts. I have found 8 similar occurrences in BNC2014 (Love et al. 2017) by 6 different speakers, all male, in four different conversations. However, the examples suggest that *son* is not necessarily used as a kinship term – the preceding speaker is in some cases of similar age or even female in two cases, e.g. you what **son?** addressed to the same age speaker.

my Mum, his Mum or *Will's mum*. *Mum and Dad*, linguistically and semantically, behave like a unit, two inseparable characters, see the following examples:

- (3) From somewhere beyond the cove came a long, eerie cry, the sound they had heard during the night. He shivered and turned to **Mum and Dad**.
[Tim Bowler, *Apocalypse*, 2012]
- (4) She closed her eyes and thought of **Mum and Dad** and Tom.
[Tim Bowler, *Buried Thunder*, 2011]
- (5) This is really not good. What will **Mum and Dad** say when we get arrested? For shoplifting!?
[Ali Sparkes, *SWITCH 10: Gecko Gladiator*, 2012]
- (6) He was glad **Mum and Dad** were there, just as he was glad Beth was there.
[Tim Bowler, *Bloodchild*, 2008]

The binomial in these examples refers to the characters of *mum* and *dad*. *Mum and dad* are two people, who as character types frequently occur in the stories and are important in the lives of the characters, as in example (4) where the main character has *Mum*, *Dad* and brother *Tom*. This is different for, e.g., ‘brothers’ and ‘sisters’:

- (7) “What are the odds of a **brother and sister** both being supply teachers and both at the same school?”
[Geraldine McCaughrean, *El Cid*, 2002]
- (8) And don’t forget, we are supposed to be **brother and sister**.
[Julia Golding, *Dragonfly*, 2008]
- (9) We were just having some **brother and sister** quality time, weren’t we, sis?
[Joss Stirling, *Stealing Phoenix*, 2011]

The stylistic differences in these examples are very subtle. However, it seems that when the binomial *brother and sister* is used, as in the above Examples (7 to 9) and other occurrences in the corpora, the semantics, the family relationship, is the reason why the particular binomial was used, see example (7), it is an unlikely coincidence that two siblings should meet in one school as teachers, or, in example (8), a certain type of behaviour is expected between brothers and sisters.

To make this claim more robust, not a mere observation, would require a more extensive study. However, it seems plausible as a contributing factor to explaining why only some kinship terms undergo the reversal. To explain the reversal itself, Tachihara and Goldberg’s (2019) cross-linguistic study of the ordering of proper names seems to be particularly relevant in this context. They conclude that both for English and Japanese speakers cognitive accessibility is the most important factor in the name ordering. So, the reversal seems to be influenced by both accessibility – which may be linked to the nature of discourse and the degree of its formality – and conceptual markedness, possibly also frequency of occurrence in fictional and other “real world” discourses, particularly spoken language, influencing each other. The fact that the

degree of informality in discourse is relevant also for the ordering is confirmed by the results in Czech, where the informal and diminutive variants are more frequent in the female-first ordering while the formal variants tend to occur in male-first ordering. On a note of caution, I should point out that the comparison of English and Czech may not be extendable in a straightforward way to other languages. Google N-gram viewer for Italian shows a similar trend with *mamma e papà* having become the more frequent variant but, for example, French does not confirm the trend, *papa et maman* still seeming to be the norm, see Figures 5 and 6.

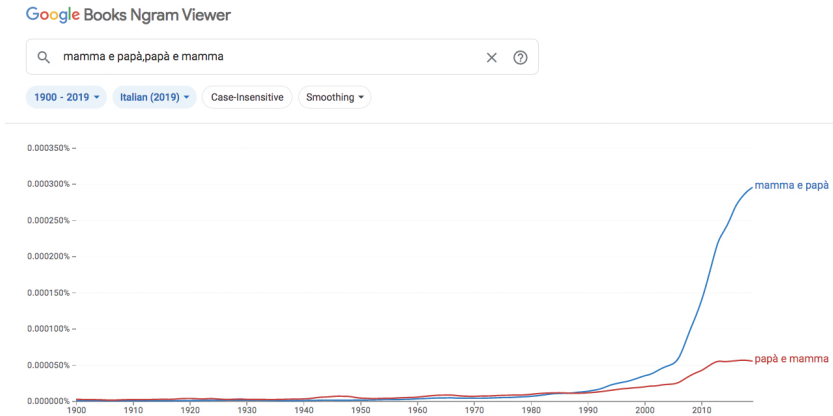


Figure 5. Google Ngram Viewer for *mamma e papà* and *papà e mamma* in Italian (accessed August 4, 2020)

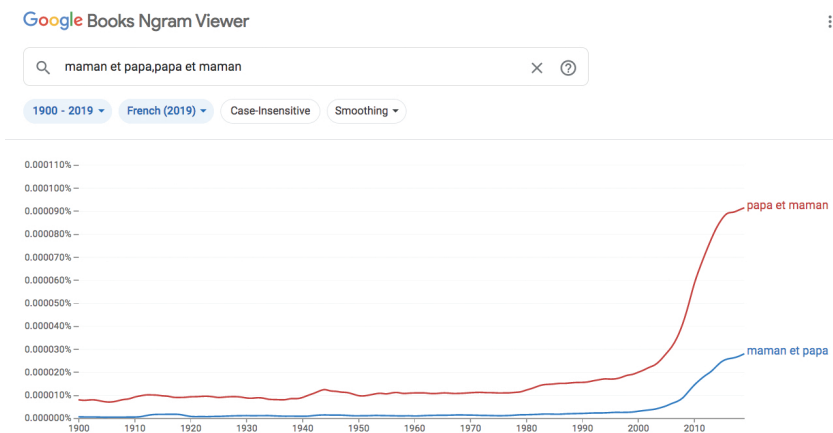


Figure 6. Google Ngram Viewer for *papa et maman* and *maman et papa* in French (accessed August 4, 2020)

7. Conclusions

This chapter addressed cross-linguistically (English and Czech) and diachronically some gaps in the research on gendered binomials. Both cross-linguistic and diachronic perspectives are, perhaps surprisingly, under-researched areas in studies on binomials. Gendered binomials have received plenty of attention in English and it has been noted that some English gendered kinship binomials have undergone a reversibility change and in the contemporary language prefer the female-first ordering, which is within the wider context of binomials the marked form. This study focuses particularly on gendered kinship binomials while using data evidence from children's fiction which constitutes one of the text-types most relevant for the study of kinship terms due to their high frequency of occurrence.

For English, the data analysis has confirmed the reversal in the nearly frozen binomial *father and mother* in the 19th century to the nearly frozen *mother and father* in contemporary language. A similar trend can be observed for the binomial *uncle and aunt*. The contemporary data also reveals the dramatic emergence of a synonymic, less formal, form *Mum and Dad*, which is in the contemporary data a frozen form with female-first ordering. Other less formal kinship binomials referring to grandparents also emerge in the female-first ordering. Whether a form is frozen or indicating an unfreezing trend is calculated through irreversibility scores. These scores are generally higher for English than Czech, which is very likely due to language typology. Czech has a “free” word order and seems to have a lower proportion of formulaic language in general.

In Czech, the results are in many ways similar. The unfreezing trend away from male-first variant is noticeable for binomials referring to parents, grandparents and ‘aunts and uncles’. However, the degree of formality also seems to be correlated, the more formal variants for parents and ‘uncles and aunts’ preferring the male-first ordering while the less formal ones prefer female-first ordering. Thus, the most applicable ordering constraints for both languages seem to be ‘perceptual markedness’ and ‘accessibility’. Morgan and Levy (2016) suggest that prior knowledge and experience with a specific binomial has an influence on the way it is expressed. In this sense, the ‘mother’ figure (and ‘aunt’ and ‘grandmother’) is actually the **unmarked** of the two (see, e.g., Craig 2016) – the language thus reflects the experience (Croft 2003). The reflection may function in two ways – subconsciously, our experience as speakers, writers and readers is shaped from early childhood, but we can also possibly consider the adjustments we make to the discourse directed at children to make it more accessible. The language in children's literature and how it differs from the language in “adult” literature is still an under-explored area.

The findings also suggest that assigning kinship terms to one and the same semantic category needs adjustments. In both languages, the reversal clearly holds only for some of the kinship binomials. Others, such as *brother and sister*, are unaffected. The reason proposed is that textually the words referring to parents, grandparents and aunts and uncles are inherently different than words referring to siblings or children. Grammatically they behave like a transitional category between common nouns and proper names. Therefore, research on proper name binomials, which shows that ‘accessibility’ is the most relevant constraint (Tachihara & Goldberg 2019), is highly relevant. However, it is quite clear that these linguistic processes are complex with a number of interlinked influences, some of which may have not been considered here.

Acknowledgements

This research was supported by the European Regional Development Fund project ‘Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World’ (reg. no.: CZ. 02.1.01/0.0/0.0/16_019/0000734) and by the programme ‘Progres Q08 Czech National Corpus’ implemented at the Faculty of Arts, Charles University. I would also like to thank two anonymous reviewers for their valuable comments and suggestions.

References

- Abrahams, Richard D. 1950. Fixed order of coordinates: A study in comparative lexicography. *The Modern Language Journal* 34(4): 276–287.
<https://doi.org/10.1111/j.1540-4781.1950.tb06011.x>
- Benor, Sarah Bunin & Levy, Roger. 2006. The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82(2): 233–277. <https://doi.org/10.1353/lan.2006.0077>
- Čermák, František. 2010. Binomials: Their nature in Czech and in general. In *Phraseologie global – areal – regional. Akten der Konferenz EUROPHRAS 2008*, Jarmo Korhonen, Wolfgang Mieder, Elisabeth Piirainen & Rosa Piñel (eds), 309–315. Tübingen: Narr.
- Čermáková, Anna. 2017. *The GLARE 19th Century Children’s Literature Corpus in CLiC* [Blog post]. <<https://blog.bham.ac.uk/clic-dickens/2017/11/28/the-glare-19th-century-childrens-literature-corpus-in-clic/>> (1 October 2019).
- Čermáková, Anna. 2018. *ChiLit: The GLARE 19th Century Children’s Literature Corpus in CLiC* [Blog post]. <<https://blog.bham.ac.uk/glareproject/2018/02/14/chilit-the-glare-19th-century-childrens-literature-corpus-in-clic/>> (1 October 2019).
- Čermáková, Anna & Chlumská, Lucie. 2017. Expressing place in children’s literature. Testing the limits of the n-gram method in contrastive linguistics. In *Cross-linguistic Correspondences. From Lexis to Genre* [Studies in Language Companion Series 191], Thomas Egan & Hildegunn Dirdal (eds), 75–96. Amsterdam: John Benjamins.
<https://doi.org/10.1075/slcs.191.03cer>

- Chaloupka, Otakar & Voráček, Jaroslav. 1979. *Kontury české literatury pro děti a mládež*. Praha: Albatros.
- Cooper, William E. & Ross, John Robert. 1975. World order. In *Functionalism*, Robin E. Grossman, L. James San & Timothy J. Vance (eds), 63–111. Chicago IL: Chicago Linguistic Society.
- Craig, Lyn. 2016. *Contemporary Motherhood: The Impact of Children on Adult Time*. London: Routledge. <https://doi.org/10.4324/9781315573878>
- Croft, William. 2003. *Typology and Universals*. Cambridge: CUP.
- Danet, Brenda. 1985. Legal discourse. In *Handbook of Discourse Analysis*, Teun A. Van Dijk (ed.), 273–291. London: Academia Press.
- Dant, Doris R. 2013. *Mom and Dad* but *Men and Women*: The sequencing of sex-determined noun pairs in American English. In *Corpus Perspectives on Patterns of Lexis* [Studies in Corpus Linguistics 57], Hilde Hasselgård, Jarle Ebeling & Signe Oksefjell Ebeling (eds), 35–46. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.57.05dan>
- Ebeling, Jarle & Ebeling, Signe Oksefjell. 2013. *Patterns in Contrast* [Studies in Corpus Linguistics 58], Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.58>
- Giammarresi, Salvatore. 2010. Formulaicity and translation: A cross-corpora analysis of English formulaic binomials and their Italian translations. In *Perspectives on Formulaic Language: Acquisition and Communications*, David Wood (ed.), 257–74. London: Bloomsbury.
- Gustafsson, Marita. 1975. *Binomial Expressions in Present-Day English. A Syntactic and Semantic Study*. Turku: Turun Yliopisto.
- Hegarty, Peter, Watson, Nila, Fletcher, Laura & McQueen, Grant. 2011. When gentlemen are first and ladies are last: Effects of gender stereotypes on the order of romantic partners' names. *British Journal of Social Psychology* 50(1): 21–35. <https://doi.org/10.1348/014466610X486347>
- Hiltunen, Risto. 1990. *Chapters on Legal English: Aspects Past and Present of the Language of the Law*. Helsinki: Suomalainen Tiedeakatemia.
- Iliev, Rumen & Smirnova, Anastasia. 2016. Revealing word order: Using serial position in binomials to predict properties of the speaker. *Journal of Psycholinguistic Research* 45(2): 205–235. <https://doi.org/10.1007/s10936-014-9341-3>
- Jerrim, John & Moss, Gemma. 2019. The link between fiction and teenagers' reading skills: International evidence from the OECD PISA study. *British Educational Research Journal* 45(1): 181–200. <https://doi.org/10.1002/berj.3498>
- Jespersen, Otto. 1905. *Growth and Structure of the English Language*. Leipzig: Teubner.
- Klégr, Aleš & Čermák, Jan. 2008. Binomials in an historical English literary perspective: Shakespeare, Chaucer, Beowulf. In *Shakespeare between the Middle Ages and Modernity: From Translator's Art to Academic Discourse. A Tribute to Professor Martin Hilský*, Aleš Klégr, Jan Čermák & Martin Procházka (eds). Prague: Charles University.
- Kopaczky, Joanna & Sauer, Hans. 2017. *Binomials in the History of English. Fixed and Flexible*. Cambridge: CUP.
- Lévi-Strauss, Claude. 1949. *Les Structures élémentaires de la parenté*. Paris: Presses universitaires de France.
- Lohmann, Arne. 2014. *English Coordinate Constructions: A Processing Perspective on Constituent Order*. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139644273>
- Love, Robbie, Demby, Claire, Hardie, Andrew, Brezina, Vaclav & McEnery, Tony. 2017. The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics* 22(3): 319–344. <https://doi.org/10.1075/ijcl.22.3.02lov>

- Lohmann, Arne & Takada, Tayo. 2014. Order in NP conjuncts in spoken English and Japanese. *Lingua* 152: 48–64. <https://doi.org/10.1016/j.lingua.2014.09.011>
- Malkiel, Yakov. 1959. Studies in irreversible binomials. *Lingua* 8: 113–160. [https://doi.org/10.1016/0024-3841\(59\)90018-X](https://doi.org/10.1016/0024-3841(59)90018-X)
- Mollin, Sandra. 2014. *The (Ir)reversibility of English Binomials* [Studies in Corpus Linguistics 64]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.64>
- Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English*. Oxford: Clarendon.
- Morgan, Emily & Levy, Roger. 2016. Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition* 157: 384–402. <https://doi.org/10.1016/j.cognition.2016.09.011>
- Motschenbacher, Heiko. 2013. Gentlemen before ladies? A corpus-based study of conjunct order in personal binomials. *Journal of English Linguistics* 41(3): 212–42. <https://doi.org/10.1177/0075424213489993>
- Nikolajeva, Maria. 2014. *Reading for Learning: Cognitive Approaches to Children's Literature* [Children's Literature, Culture, and Cognition 3]. Amsterdam: John Benjamins. <https://doi.org/10.1075/clcc.3>
- Onishi, Kristine H., Murphy, Gregory L. & Bock, Kathryn. 2008. Prototypicality in sentence production. *Cognitive Psychology* 56(2): 103–141. <https://doi.org/10.1016/j.cogpsych.2007.04.001>
- Pordány, László. 1986. Irreversible binomials in Hungarian. *Acta Linguistica Academiae Scientiarum Hungaricae* 36(1-4): 151–79.
- Reynolds, Kim. 2011. *Children's Literature: A Very Short Introduction*. Oxford: OUP. <https://doi.org/10.1093/actrade/9780199560240.001.0001>
- Siyanova-Chantura, Anna, Conklin, Kathy & van Heuven, Walter J. B. 2011. Seeing a phrase time and again matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition* 37(3): 776–784.
- Sobkowiak, Włodzimierz. 1993. Unmarked-before-marked as a freezing principle. *Language and Speech* 36(4): 393–414. <https://doi.org/10.1177/002383099303600403>
- Tachihara, Karina & Goldberg, Adele E. 2019. Cognitive accessibility predicts word order of couples' names in English and Japanese. *Cognitive Linguistics*. <https://doi.org/10.1515/cog-2019-0031>
- Wright, Sandra & Hay, Jennifer. 2002. Fred and Wilma: A phonological conspiracy. In *Gendered Practices in Language*, Sarah Benor, Mary Rose, Devyani Sharma, Julie Sweetland & Quing Zhang (eds), 175–91. Stanford CA: CSLI.
- Wright, Sandra K., Hay, Jennifer & Bent, Tessa. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43(3): 531–561. <https://doi.org/10.1515/ling.2005.43.3.531>

Tracing processes in auxiliarization

Time-sufficiency verbs from a Norwegian-Swedish-English contrastive perspective

Mats Johansson & Lene Nordrum

Lund University

This chapter considers the Norwegian verb *rekke* ('reach') and its English correspondences in a bi-directional translation corpus (ENPC), as well as occurrences of *rekke* in Norwegian monolingual corpora. We suggest that *rekke* is undergoing grammaticalization from a concrete lexical spatial verb towards a (semi-)auxiliary that contains a component of time sufficiency, similar to Swedish *hinna* (Johansson & Nordrum 2016), but that *rekke* is at an earlier stage in this development. We base this suggestion on the following observations: (1) Swedish *hinna* was earlier polysemous between space and time, but can only denote time in present-day Swedish. (2) Norwegian *rekke* is still polysemous between space and time, but seems to always denote time when used as a (semi-)auxiliary. The study contributes to the semantic description of Norwegian *rekke*, but also has implications for grammaticalization studies as well as for studies of modality.

1. Introduction

This chapter focuses on the Norwegian verb *rekke* ('reach') viewed both through its English correspondences in the *English-Norwegian Parallel Corpus* (ENPC) and in Norwegian monolingual corpus data, and compares it to the Swedish verb *hinna* (roughly: 'have enough time to be able to [do sth]'). The study continues previous work exploring the semantics of a group of verbs referred to as sufficiency verbs (Flint 1980; Johansson & Nordrum 2016, 2018), i.e. verbs that involve a notion of there having to be a sufficient amount of some entity at disposal for an action to be possible. The English translation of Norwegian *rekke* in (1) illustrates how *rekke* can be used to denote whether there is sufficient time for some event to take place.

- (1) Bruno hoiet ut gjennom vinduet til en pike som vasket tøy på gårdsplassen
 før jeg *rakk* å protestere. (SL1)
Before I had time to protest Bruno called from the window to a girl washing
 clothes in the yard. (SL1T)¹

In (1), the translation of *rekke* with ‘had time’ captures the meaning that sufficient time is necessary for the subject (*I*) to be able to protest. In this chapter, we argue that this is the only interpretation possible for *rekke* in (1), although it can have other meanings in other contexts. Further, we propose that such uses express modality, and thus fit into the paradigm of Norwegian modal verbs (see Eide 2005). Our reasoning here is that if we define modality in line with van der Auwera and Plungian (1998) and von Stechow (2006), namely as a concept that refers to dimensions of possibility or necessity, time sufficiency falls under van der Auwera and Plungian’s category of participant-external possibility.

Although sufficiency as a concept has gained a great deal of attention in the literature (Fortuin 2013), especially in approaches to adverbs such as *enough*, sufficiency verbs have largely escaped scholarly attention. Exceptions include Flint (1980), who lists 45 sufficiency verbs in Finnish, and Johansson and Nordrum (2016, 2018), who discuss the Swedish verbs *hinna* (Johansson & Nordrum 2016) and *orka* (Johansson & Nordrum 2018). Presumably, the main reason for the lack of attention is that sufficiency verbs are typologically rare. They occur in the languages around the Baltic sea (Flint 1980; Johansson & Nordrum 2016, 2018; Nadathur 2016), but to our knowledge, they are infrequent outside of this region. Sufficiency verbs are thus interesting both from a contrastive point of view, since they lack straightforward equivalents in other languages, and in their own right, by virtue of being typologically rare.

Johansson and Nordrum (2016) use English correspondences in the *English-Swedish Parallel Corpus* (ESPC) to shed light on the semantics of *hinna*, a Swedish sufficiency verb meaning roughly ‘have enough time to be able to [do sth]’, and show that it encodes two meaning components: time sufficiency as its core, at-issue sense, and ability as presupposition. The semantics is made visible in the English correspondences of *hinna* in the ESPC, which always include temporal expressions, reflecting the at-issue sense, and often, but not always, combine these with ability verbs like *can*, reflecting presupposed ability.

1. In all of the examples from ENPC, translations are identified by a capital T at the end of the reference to the corpus text. For a full list of the corpus texts, see <https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/> (30 May 2020)

Our aim in this chapter is to view *rekke* through its English correspondences in the ENPC in a similar way as for Swedish *hinna*, and compare the results with the account of *hinna*. We pose the following research questions:

1. Which meanings of *rekke* are reflected in the English correspondences in the ENPC?
2. Specifically, does the parallel data support our intuition that *rekke* can have a similar meaning as *hinna*, i.e. as a time-sufficiency verb expressing participant-external possibility (i.e. modality)?
3. If there are different meanings: Are there any clear lexical correspondences reflecting different meanings of *rekke* and/or any form-meaning patterns, and how do they differ from *hinna*?
4. If there are form-meaning patterns: Do these patterns show up in monolingual corpora as well, and if so, how can they be explained?

Importantly, despite forming the background of our study, the concepts of sufficiency and modality do not serve as the starting point for our analysis. Rather, we follow Chesterman (1998) in regarding the *tertium comparationis* of translations as the result of the analysis rather than its basis. Put in traditional contrastive analysis terms, we approach the parallel corpus data from the point of view of the verb *rekke*, rather than from the point of view of a *tertium comparationis* represented by the concept time sufficiency, i.e. an independent meaning component forming the point of departure for a comparison of Norwegian and English expressions.

This chapter is organized as follows. Section 2 provides some initial observations about Norwegian *rekke* in contrast with Swedish *hinna*. Then follows a description of the general procedure of the study and the corpora used (Section 3). Section 4 presents the correspondences of *rekke* in the parallel corpus and also compares the correspondences of *rekke* to those found for *hinna* in Johansson and Nordrum (2016). The patterns found in the parallel corpus data are then explored in monolingual corpora in Section 5. Section 6 proposes that based on the results presented in Sections 4 and 5, *rekke* is undergoing grammaticalization and becoming more like *hinna*, i.e. a modal semi-auxiliary denoting time sufficiency. Section 7 provides a summary of our conclusions and some final remarks.

2. Some initial observations about Norwegian *rekke* – in contrast with Swedish *hinna*

Even if Norwegian *rekke* can correspond to Swedish *hinna*, it is not the only possible correspondence. For example, one alternative candidate mentioned by Swedish-Norwegian dictionaries is the verb *nå* (see e.g. *Svensk-norsk blå ordbok*

2014: 162).² There are at least two reasons for limiting ourselves to a study of *rekke*, however. The first is that *rekke* appears in so-called ‘parallel translations’ (Hasselgård 2006; Johansson 2007; Nordrum 2015) in the sister corpora ENPC and ESPC, meaning that *hinna* and *rekke* are used in Swedish and Norwegian translations of the same English source material, as illustrated in (2).

- (2) *Before he could add*, as he had begun to, suppressing a tone of irony, “Only the people”, she exclaimed, “Thank God for that!” (RR1)

Han hade tänkt säga “bara människorna”, men utan att låta alltför ironisk, men *hann inte förrän hon utropade*: “Tack och lov för det!” (RR1T)

Før han rakk å tilføye, i en tilbakeholdt ironisk tone: “Bare menneskene,” utbrøt hun: “Gud skje takk og lov!” (RR1T)

The second reason is that *rekke* is versatile enough for its own study. The Norwegian dictionary, *Det Norske Akademis Ordbok* (NAOB),³ lists four related meanings (below), of which the fourth is described as archaic and only used in fixed expressions.

- i. (ved sin utstrekning, lengde eller rekkevidde) *nå* (roughly: ‘reach by means of size, length or reach’)

Examples:

Det hvite foldeskjørtet rakk helt ned til ankene (Lit. ‘The white pleated skirt reached completely down to the-ankles’)(Herman Willis: St. Olav, LBK, 2004)

Han hadde fulgt ordre så langt hans evner og konsentrasjon rakk (Lit. ‘He had followed orders as far as his abilities and concentration reached’)(Terje Stigen: *Ved foten av kunnskapens tre*, LBK, 1986)

- ii. *berøre* (og få tak i) *ved å strekke seg* (roughly ‘touch (and get hold of) by means of reaching out (extending oneself)’)

Example:

Han kan ikke række øksen (Lit. ‘He cannot reach the-axe’) (Knut Hamsun: Markens Grøde II, 45, 1917)

- iii. *nå frem til* (sted eller transportmiddel) *tidsnok* (roughly: ‘reach a place or means of transport in time’)

Example:

Vi skal ... rekke en buss (Lit. ‘We shall ...reach a bus’)(Ingvild H. Rishøi: *Vinternoveller* 80, 2014)

2. In Danish, *nå* seems to be the verb that most closely corresponds to Swedish *hinna* (*Ordbok over det Danske Sprog* –<https://ordnet.dk/ods>, (30 May 2020).

3. NAOB https://www.naob.no/ordbok/rekke_5 (25 November 2019).

iv. *vedvare* (roughly: ‘persist’)

Example

Den ligevægt, som rækker gennem årene (Lit. ‘The equilibrium, which reaches through the-years’)(Henrik Ibsen: *Kærlighedens komedie* 145, 1873)

Notably, however, the entries in NAOB do not support our intuition that *rekke* is like *hinna* (as suggested by the parallel translations in (2)). For one, unlike *hinna*, *rekke* seems to be polysemous between space and time. Entries (i) and (ii) denote space meanings and can be paraphrased as ‘something/somebody reaches somewhere’, while entries (iii) and (iv) reflect time meanings, in the case of (iii) ‘something/somebody getting somewhere in time’ and in (iv), something staying the same way through time. Secondly, unlike *hinna*, which is most commonly used as a (semi-)auxiliary (Johansson & Nordrum 2016), all the examples of *rekke* in NAOB involve main-verb uses. These observations raise the question whether examples such as (1) and (2) above, where *rekke* is used with an *å*-infinitive complement and expresses time sufficiency, are in fact marginal.

3. The general procedure of the study and corpus data

The study of *rekke* involved four types of corpora that were investigated in three, partly intertwined, steps. First, following the general methodology of “seeing language through multilingual corpora” (Johansson 2007), *rekke* was viewed through English translation correspondences in the ENPC, a bidirectional translation corpus of approximately 2.6 million words with original texts in both English and Norwegian and translations into the other language. The instances of *rekke* and their English correspondences were retrieved from the sentence alignment interface for the ENPC, transferred to a database, and then annotated for syntactic context and meaning.

Second, the results from the translation data were checked against Norwegian monolingual corpora of written and spoken language, where random instances of *rekke* were retrieved and coded for syntactic context and meaning. The monolingual corpora of written language comprise *Leksikografisk Bokmålskorpus* (LBK) (‘Lexicographic corpus for Norwegian Bokmål’), containing about 100 million words (Knudsen & Vatvedt Fjeld 2013), the *Norwegian Web as a Corpus* (NoWac v. 1.0), containing about 700 million words (Guevara 2010), and *Norwegian Web 2017* (NoTenTen17, Bokmål), available via the search interface Sketch Engine. For spoken material, we used *Norsk Talespråkskorpus, Oslodelen* (NoTa-Oslo) (‘Norwegian Corpus of spoken language, the Oslo part’) which comprises 957 000 words (Johannessen & Hauge 2008).

In a third step, partly overlapping with steps one and two, the patterns and hypotheses generated by the study of *rekke* were compared with the analysis of

Swedish *hinna* in Johansson and Nordrum (2016). This comparison involved supplementary data from the Swedish monolingual corpora available via *Språkbanken* ('Swedish Language Bank'), University of Gothenburg (Borin et al. 2012).

4. *Rekke* in the ENPC compared to *hinna* in the ESPC

This section presents the correspondences of *rekke* in the parallel corpus and compares them to the results for *hinna* in Johansson and Nordrum (2016). We begin with a short summary of the results for *hinna* before we go on to consider the English correspondences of *rekke*.

4.1 Swedish *hinna* revisited

For *hinna*, Johansson and Nordrum (2016) show that time-sufficiency meaning is the core, or 'at-issue', meaning in combination with presupposed ability. A couple of observations are crucial to this conclusion. First, all the English correspondences of *hinna* involve a time-expression, either as a congruent translation, or indicated by a temporal element in the context. Example (3) gives an example where a time expression is added in the English translation (underlined) (Johansson & Nordrum 2016: 186).

- (3) Jag ser ronden komma men **hinner** inte varna henne. (PCJ1)
I see them coming but don't **manage** to warn her in time. (PCJ1T)

In (3), *hinna* is translated by the ability verb *manage*, while *in time* is added to capture the time-sufficiency meaning. Second, when a time expression has a congruent correspondence, there is no overt expression of ability in the English correspondence, as illustrated in (4) (Johansson & Nordrum 2016: 178).⁴

- (4) Hon hade inte **hunnit** bli riktigt vuxen när hon fick mig. (AP1)
She hadn't really **had time** to become grown up when she had had me. (AP1T)

It should be pointed out, though, that the study of *hinna* only involves (semi-)auxiliary uses, which account for almost all of the examples (Johansson & Nordrum 2016). Some examples were found, however, of main-verb uses followed by an adverbial, which correspond to English expressions with motion verbs such as *get*. Example (5) is an illustration (adverbial underlined) (Johansson & Nordrum 2016: 176).

4. It is quite possible, however, that covert ability is expressed in (4) by means of the infinitive clause *to become grown up* (see Bhatt 1999).

- (5) Redan innan han hade **hunnit fram** började han tala lugnande till sig själv,
inåtvänt och utan att röra på läpparna. (KOB1T)
Even before he **got there**, he had started talking to himself – Inwardly,
without moving his lips, in order to cool down. (KOB1T)

On account of their scarcity, examples such as (5) were not included by Johansson and Nordrum (2016), but it was speculated that they reflect an historically older and non-modal meaning of *hinna* close to its etymological sources Gothic *hinþan* ('catch') and Old High German *heriunda* ('prey') as well as English *hunt*, and Old Norse *inna* ('reach an end', 'finish') (see *Svenska Akademiens Ordbok* (SAOB)). We return to this observation in our analysis of *rekke*.

4.2 The English correspondences of *rekke* at a glance

As a first step, we consider the English correspondences of Norwegian *rekke* in the ENPC. They are listed in Table 1 and classified in terms of fairly broad meaning categories.⁵

Table 1. *Rekke* and its English correspondences in the ENPC⁶

English correspondence	NO→ET	NT←EO
	<i>N</i> = 52	<i>N</i> = 27
<i>Get</i> -correspondences <i>reach/get to/catch/stretch/return</i>	17	12
<i>Ability</i> -correspondences <i>be able to/can/manage/have a chance</i>	16	9
<i>Time</i> -correspondences <i>take the time/have time/in time/already</i>	6	2
<i>try/start</i>	2	-
<i>zero</i>	2	-
<i>be enough</i>	1	-
<i>other</i>	8	4
Total	52	27

5. We regard *manage* as denoting actualized ability (see Johansson & Nordrum 2016, Nadathur 2016).

6. The term 'correspondence' refers to both the English translations (NO→ET) and the sources (NT←EO) of *rekke* (Johansson 2007: 23). The table includes both congruent correspondences and divergent correspondences. In congruent correspondences, the source syntax is kept intact (Johansson 2007: 23–26), i.e. *rekke* corresponds to a verb in English and the complement structure is the same, while divergent correspondences have a different syntactic structure than the original, as in the correspondence *be enough*. None of the correspondences listed in Table 1 involve major syntactic changes.

Some similarities and differences with the results for *hinna* (Johansson & Nordrum 2016) can be gleaned from Table 1. Similarly to *hinna*, ability correspondences (e.g. *be able to*, *manage*) and time correspondences (e.g. *have time*, *in time*) show up as two major categories. However, unlike *hinna*, the most frequent English correspondence is the group of motion verbs (*reach/get to/catch/stretch/return*), henceforth *get*-correspondences. Also, different from *hinna* is the translation with *be enough*, where a closer look suggests that *rekke* denotes ‘general sufficiency’: In this case, whether someone has sufficient money to last for the duration of a journey, illustrated in (6).

- (6) Det som stanset ham var at han nesten ikke hadde penger igjen,
og ikke trodde de ville **rekke** den lange veien hjem til Wien. (EFH1)
What had stopped him was that he had hardly any money left and didn’t
think it would **be enough** for the long journey back to Vienna. (EFH1T)

In what follows, we discuss the three most common translation categories, leaving aside the ‘general sufficiency’ correspondence (in 6) on account of low numbers. We first discuss the English correspondence types that are similar to *hinna*, ability and time, and then move on to a discussion of the *get*-correspondences.

4.2.1 *Rekke with ability correspondences*

The second-most frequent translation correspondence of *rekke* (see Table 1) is an ability correspondence such as *can*, *be able to* or *manage* (recall that we understand *manage* to denote actualized ability). At face-value, this suggests that *rekke* primarily encodes ability, but as for *hinna*, a qualitative look at the examples reveals that “pure” ability is not at stake, but rather the ability of completing an action/event within a contextually given time-frame, typically expressed in the form of a temporal “limiter” such as a temporal clause (8 of 16 NO→ET correspondences and 6 of 9 of NT←EO correspondences in Table 1 contain a temporal clause). This context is illustrated in (7) and (8) (*rekke* and its ability correspondences in bold, temporal clause in italics).

- (7) Og før han **rakk** å tenke tanken ut, hørte han en menneskerøst som
gjorde ham stivklaka og nummen fra halsen og ned til føttene: [...] (KAL1)
*Before he **could** complete his thought*, he heard a voice that scared
him stiff and made him numb from head to toe: [...] (KAL1T)
- (8) *Before I **could** do anything* the women bundled me in, clambered on,
and rowed us across to an island that wasn’t far away. (BO1)
*Før jeg **rakk** å gjøre noe*, fikk kvinnene meg ombord, kløv inn og
rodde oss over til en øy som ikke lå langt unna. (BO1T)

The temporal context can also be vaguer, as in (9), where the time it takes to finish one event, that of ironing a pile of clothes, conditions the possibility of finishing another event, that of preparing dinners in advance.

- (9) Haugen av stryketøy er like stor, hvis hun legger bort resten til imorgen, **rekker** hun ikke å lage opp middag for et par dager, slik hun gjør hver onsdag. (BV2)
 The pile of ironing is as high as ever. If she leaves the rest till tomorrow then she will not **be able to** prepare dinner for a couple of days in advance, as she does every Wednesday. (BV2T)

Our data only include three instances of *rekke* with ability correspondences where Swedish *hinna* would not be possible.

- (10) Ætten hans kom fra Hvaler og hadde vært knyttet til sjøen
så langt tilbake minnet rakk. (KH1)
 His family came from Hvaler and had been connected with the sea
as far back as they could remember. (KH1T)
- (11) Ser bare lekke glattslikka hus *så langt øyet rekker*, selvtilfredsheten
 skummer rundt dem. (BV2)
 All I notice is sleek tasteful houses *as far as the eye can see*, waves
 of self-satisfaction lapping around them. (BV2T)
- (12) Langt mot øst og vest strakte det tørre slettelandet seg *så langt øyet rakk.*
 (SH1)
 Far away to the east and west the dry prairies stretched out *as far as the eye*
could see. (SH1T)

In (10) to (12), there is no overt time expression. Instead, we get a combination of *rekke* and a spatial adverbial *så langt* ('as far as'), suggesting that ability correspondences can reflect both time and space meanings of *rekke*, and that the context is necessary for disambiguation.⁷

4.2.2 *Rekke with time correspondences*

All the time correspondences of *rekke* (e.g. *have time* or *take time*, see Table 1) reflect a time-sufficiency meaning, like *have time* in (13) and (14).

- (13) Bruno hoiet ut gjennom vinduet til en pike som vasket tøy på gårdsplassen
 før jeg **rakk** å protestere. (SL1)

7. The adverbial *så langt* ('as far as') in (10) may reflect a time rather than space meaning on account of metaphorical/metonymic relations. In (10)-(12), a metaphoric/metonymic entity functions as the subject in the Norwegian source (*minne* ('memory') in (10), and *øye* ('eye') in (11) and (12)), and the English translations feature a verb denoting the process to which the metaphoric/metonymic entities are intimately related: For *minnet rakk* ('memory reached'), we get *can remember* and an added subject, *I*, and for *øyet rekker/rakk* ('the eye reaches/reached'), *eye* is kept as a metonymic entity and *rakk* is translated with *can/could see*.

Before I **had time** to protest Bruno called from the window to a girl washing clothes in the yard. (SL1T)

- (14) **Have I time** for me to finish feeding Deirdre and Cathy before we put it on? (RDO1)

Rekker jeg å gjøre meg ferdig med å mate Deirdre og Cathy først? (RDO1T)

We also note that, similarly to the ability correspondences, the time correspondences typically involve a temporal expression in the immediate context, rendering a sort of “double” time meaning, and that ability meaning is presupposed. Support for presupposed ability in (13) is that the subject (*I*) needs to have the ability to protest in order for the possibility of protesting to be denied.

4.4.3 *Rekke with å-infinitive complementation vs. other types of complements*

In contrast to the results for *hinna*, where the majority of the corpus examples are semi-auxiliary uses (Johansson & Nordrum 2016), the corpus examples of *rekke* considered so far involve both *å*-infinitive complements and other types of complements typical of main verbs. We have already noted three main verb uses of *rekke* (see (10)-(12) above), and there are more. In four Norwegian sources with ability translations, *rekke* is followed by an NP complement combined with a time-expression, as illustrated in (15) (*rekke* and ability correspondence in bold, complements underlined, time-expression in italics).

- (15) Kwart på fire slipper de ut, hun **rekker** noen innkjøp *før de setter seg i bilen*. (BV2)

At a quarter to four they are let out, and she **manages** to do a bit of shopping *before they get into the car*. (BV2T)

Further, for the time correspondences, one of six Norwegian sources and one of two Norwegian translations involve *rekke* as a main verb.

At this point, we can conclude that a closer look at the contexts of the ability and time correspondences of *rekke* suggests that its meaning is similar to that of *hinna*, but also different. Similar to *hinna*, the bulk of the examples that correspond to time and ability translations denote time sufficiency with presupposed ability. As regards differences, we see that some of the ability translations reflect *rekke* with a space meaning and that *rekke* is frequent as a main verb.

4.2.4 *Get-correspondences*

We now turn to the most frequent category of English correspondences in Table 1, the *get*-correspondences, illustrated in (16).

- (16) Tora innbilte seg at dersom “huset” ikke hadde fått klippet håret av mora, så hadde det **rukket** helt ned til hoftene. (HW1)

Tora imagined that if the house hadn't cut off her mother's hair, it would have **reached** all the way down to her hips. (HW1T)

In (16) *reach* clearly translates main verb *rekke* with spatial meaning. It describes the length of someone's hair in terms of 'how far' the hair reaches down on the referent's body. The example mirrors entry (1) in NAOB (see Section 2), and thus underlines that *rekke* can have unambiguous space meanings.

Not all the *get*-correspondences reflect space meanings, however. In fact, for 9 of the 17 NO→ET correspondences and 8 of the 12 NT←EO correspondences in Table 1, a time-sufficiency interpretation is more likely. Consider (17).

- (17) Men så ser faren på klokken, det er *på tide* å **rekke** toget. (EFH1)
But then his father looks at his watch and *it's time* to **catch** the train. (EFH1T)

In (17), *rekke* is followed by an NP complement, *toget* ('the train') denoting a spatial endpoint, and there is a time expression in the context (in italics). Clearly, the endpoint of an event/activity is in question: Someone needs to be at a certain place (at the train), at a certain time (before the train leaves). The focus on the endpoint is captured in the translation with the achievement verb *catch*. Arguably, *rekke* still involves a type of space reading in (17) in that the subject participant needs to move from one place to another, but surely what is at stake is the time pressure in question, or put differently, the time sufficiency.

In light of the above, we sorted *get*-correspondences such as (17) into a separate semantic category: 'time with spatial endpoint'. We then considered the *get*-correspondences again, this time with a careful look at whether *rekke* seems to denote space, time sufficiency, or time with spatial endpoint. Table 2 presents the distribution among these categories.

Table 2. The *get*-correspondences of *rekke* in the ENPC, divided according to the meaning of *rekke*

English correspondences	Space		Time sufficiency		Time with spatial endpoint		N = 28
	NO	NT	NO	NT	NO	NT	
<i>reach/get/catch/ stretch/return</i>	8	3	1	1	8	7	28

Table 2 shows that there is an almost even split in terms of the meaning reflected: Roughly half of the *get*-correspondences (11 of 28) reflect *rekke* with space meaning and half (15 of 28) reflect *rekke* with time with spatial endpoint meaning. In fact, a closer look at the *get*-examples in each of the meaning categories yields a clear division of labor among the verbs: *get*, *catch*, and *return* encode time with spatial endpoint, whereas *reach* and *stretch* encode space. Based on this pattern,

we split the *get*-correspondences into two categories: *get*, *catch*, *return* on the one hand, and *reach*, *stretch* on the other. Table 3 lists the correspondences according to the three types of meaning.

Table 3. The *get*-correspondences of *rekke* in the ENPC, divided according to the meaning of *rekke*⁸

English correspondences	Space		Time sufficiency		Time with spatial endpoint		N = 78
	NO	NT	NO	NT	NO	NT	
<i>get/catch/return</i> (17)			1	1	8	7	17
<i>reach/stretch</i> (11)	8	3					11
<i>be able to/can/manage/have a chance</i> (25)	3	-	13	9	-	-	25
<i>take the time/have time/in time/already</i> (8)	-	-	5	2	1		8
<i>try/start</i> (2)	-	-	2	-			2
<i>zero</i> (2)	-	-	2	-			2
<i>be enough</i> (1)	-	-					-
<i>other</i> (13)	1	1	6	4	1		13
Total	12	4	29	16	10	7	78

As can be seen, the majority of the occurrences of *rekke* have time meaning (a total of 62 of 78, as indicated by the bold numbers in the bottom row). This distribution is somewhat surprising in light of the description of *rekke* in NAOB (see Section 2). For one, the meaning with the most instances in Table 3, time sufficiency, is not exemplified in NAOB, which only includes an example of time with spatial endpoint (see entry iii, Section 2). Further, as already pointed out, NAOB does not include examples with *å*-infinitive complementation, which is a common realization of time-sufficiency meaning in our material (see e.g. examples (7) and (8) in Section 4.2.1).

4.3 Summing up the parallel corpus data

At this point, we can summarize the results from the translation corpora in light of our research questions. Regarding research question 1, which meanings of *rekke* are reflected in the English correspondences in the parallel corpus, it is clear that time

8. The row ‘Total’ sums up the total of each semantic category, with the right-most number giving the total number of examples. The numbers in bold show that the two types of time meanings account for the bulk of the examples (62 of 78).

meanings dominate, but there are also space meanings. In addition, one translation correspondence, *be enough*, is included, which reflects a ‘general sufficiency’ meaning. Regarding research question 2, whether the parallel data provide evidence that *rekke* can have a similar meaning to *hinna*, i.e. as a time-sufficiency verb expressing participant-external possibility (modality), it is clear that *rekke* can indeed be a time-sufficiency verb, but we have not yet discussed its modal status.

Regarding research question 3, which asked whether there are any lexical or form-meaning patterns visible in the data for Norwegian *rekke*, and whether they differ from *hinna*, some patterns have been indicated:

1. Time-sufficiency meanings are typically mirrored in English by ability and time correspondences combined with a temporal context, and often correspond to *rekke* with an *å*-infinitive complement, but not always.
2. A second time meaning, time with spatial endpoint, is mirrored by the translation correspondences *get/catch/return*, reflecting *rekke* as a main verb.
3. Space meanings are typically mirrored by the correspondences *reach/stretch*, and show up for *rekke* as a main verb.
4. Beyond time and space, there is a hint (only 1 example) of a third meaning category, ‘general sufficiency’, mirrored by the correspondence *be enough*, which reflects the use of *rekke* as an intransitive main verb. Compared to *hinna*, pattern (1) is similar, but (2–4) are different.

In what follows, we address research question 4, whether the form-meaning patterns suggested in the parallel corpus data show up in monolingual corpora as well, and if so, how they can be explained. More specifically, we ask first whether the dominance of time meanings found in the ENPC is supported in large monolingual corpora, and second, whether *rekke* with an *å*-infinitive complement is in common use, and if so, if it is always associated with time sufficiency.

5. *Rekke* in monolingual corpora

To explore research question 4, we culled 200 random examples of the lemma *rekke* from two monolingual corpora: *Leksikografisk Bokmålskorpus* (LBK) and the *Norwegian Web as a Corpus* (NoWac 1.0), as well as the only 50 examples of the lemma *rekke* from the corpus *Norsk Talespråkskorpus, Oslodelen* (NoTa-Oslo). We then performed a semantic analysis of these examples. Of the two written corpora, only LBK is comparable to the ENPC in that it consists of both literary and non-literary texts published in the time span 1985–2013 (the ENPC texts are predominantly from the 1980s and 1990s). The other corpora are not comparable, but give us a broader picture of present-day Norwegian. NoWac 1.0 consists of

Norwegian texts published on the internet between November 2009 and January 2010, and has the advantage that it includes a number of texts that have not been professionally edited (including blogs and entries from discussion boards), and might thus capture language that is closer to present-day spoken Norwegian. The corpus NoTa-Oslo, lastly, includes conversations by Norwegian speakers born and raised in the greater Oslo area in Norway collected 2004–2006, and provides a glimpse of authentic spoken Norwegian.

Table 4 shows the semantic types of 200 random examples in LBK and NoWac 1.0, as well as the 50 examples the lemma *rekke* in NoTa-Oslo. The four categories in Table 4 are illustrated in (18) to (21).

Table 4. The semantic types of 200 random examples⁹ of *rekke* in LBK and NoWac 1.0 as well as 50 examples from NoTa-Oslo

Meaning	LBK <i>N</i> = 200	NoWac <i>N</i> = 200	NoTa-Oslo (<i>N</i> = 50)
Time sufficiency	138	132	32
Time with spatial endpoint	23	35	12
Space	29	26	5
General sufficiency	10	7	1
Total	200	200	50

Time sufficiency:

- (18) Igår var vi hjemme litt før åtte, så jeg **rakk** å bestille pizza på shell.
 [Lit. ‘Yesterday were we home a bit before eight, so I had time to order pizza on shell’]
 (NoWac, sentence id 1970874)

Time with spatial endpoint:

- (19) Ungdommen har innrømmet at han hadde det travelt. Han skulle **rekke** en ferje til Molde.
 [Lit. ‘The-youth has admitted that he had it stressful. He was-to reach a ferry to Molde’]
 (LBK, SA11Mo0601.2358)

Space:

- (20) [...] har vi gått ned enda en størrelse til hvit svelgtube. Denne **rekker** tydelig ikke ned til pasientens kjeveben, og er således for kort.
 [Lit. ‘[...] have we gone down yet a size to white endotracheal tube. This reaches apparently NOT down to the-patient’s jawbone, and is thus too short’]
 (NoWac, sentence id 605358)

9. The 200 random examples in LBK and NoWac were manually cleaned so that no examples came from the same texts.

General sufficiency:

- (21) Bellona har regnet ut at 10% av inntektene av CO 2 – avgiften *rekker* til et omfattende prosjekt om hydrogen-fyllestasjoner spredt over hele landet.
 [Lit. ‘Bellona has calculated out that 10% of the-income of Co 2 tax is-sufficient to a comprehensive project about hydrogen filling stations spread across whole the-country’] (NoWac, sentence id 35414831)

As shown in Table 4, the distribution of types in both corpora supports the findings from the ENPC: The majority of the examples denote a time reading, of which time sufficiency is by far the most common category (in bold) for all corpora. This shows that time sufficiency is an established meaning for *rekke*. One observation worth noting is that the ‘general sufficiency’ category indicated in the parallel corpus material by one example (*be enough*), also shows up in the monolingual data.

Importantly, the analysis of the examples in Table 4 revealed a clear form-meaning relation: In all the cases when *rekke* is used with an *å*-infinitives or bare infinitive complement in LBK (117 examples), NoWaC 1.0 (82 examples) and NoTa (12 examples), we get time-sufficiency meanings, whereas with other complements, time, space, and general-sufficiency interpretations are possible. Thus, it is clearly in its use with infinitival complements that *rekke* most closely resembles Swedish *hinna*, although we note that *hinna* can also be used with NP-complements corresponding to example (19), at least if the NP denotes a means of transportation.

6. Is *rekke* becoming grammaticalized?

We have seen that *rekke* with time-sufficiency meaning often occurs with an *å*-infinitive complement, which is similar to Swedish *hinna*. In addition, *rekke* has spatial meanings (example (20) in Section 5) and general sufficiency meanings (example (21) in Section 5) in other contexts, which is impossible for *hinna*. Interestingly, however, *hinna* had spatial meanings in the past. In the *Swedish Academy Dictionary* (*Svenska Akademiens Ordbok* (SAOB)), the following examples are noted.

- (22) Om de nu kunde förmås at allena nyttja en (kjortel), som **hunne** til knäet.
 (1769)
 [Lit. ‘If they now could be-MADE to alone use a skirt, which reached to the-knee.’]
- (23) Min syn **hinner** ej så långt.
 (1852)
 [Lit. ‘My sight reach NOT so far.’]

Examples (22) and (23) are not possible in present-day Swedish. However, the fact that *hinna* involved a space meaning in the past raises the question whether it is

possible that the two verbs have developed in similar ways, but that *rekke* is behind *hinna* in this development. More specifically, we hypothesize that:

1. Norwegian *rekke* is behind Swedish *hinna* in space-to-time semantic change: *hinna* was earlier polysemous between space and time, but can only denote time in present-day Swedish. In contrast, Norwegian *rekke* is still polysemous between space and time.
2. Swedish *hinna* is a semi-auxiliary. Norwegian *rekke* seems to be on its way to becoming grammaticalized into a semi-auxiliary, occurring both with *å*-infinitives and (less frequently) with bare infinitives.

Hypotheses 1 and 2 are possibly related, so that when *rekke* is used as a (semi-) auxiliary, meanings other than temporal ones are not possible.

These hypotheses are informed by at least two major observations from grammaticalization studies. First, semantic change by way of metaphorical extensions is typical in the early stages of grammaticalization, and a shift from SPACE to TIME has been observed across languages (Heine, Claudi & Hünemeyer 1991). For *rekke*, we hypothesize that the time meanings have developed from space meanings. Second, a change from lexical verb to auxiliary is regarded as a prime example of grammaticalization. In this process, the verb gradually loses syntagmatic variability, such as the ability to take NP complements, and becomes syntactically restricted to taking VP complements. The syntactic change typically also involves ‘semantic bleaching’ (Hopper & Traugott 1993: 87) of the lexical meaning to the advantage of grammatical meaning – a prime example is the English verb *will*, where ‘volition’ has been regarded as ‘bleached’ when it is used as an auxiliary, and the grammatical meaning ‘future’ is salient (Hopper & Traugott 1993: 92–93). For *rekke*, we note that our material includes some examples with bare-infinitive complements, a form that is regarded as unacceptable by many native speakers of Norwegian. This is an indication that *rekke* may be developing into a true auxiliary, bare-infinitive complementation being the top criterion for auxiliary status in Norwegian descriptive grammars (Faarlund et al. 1997). As an auxiliary, the lexical meaning ‘space’ would be bleached to the advantage of ‘time sufficiency’, which we understand as a modal category (see Section 6.1.2).

Observations from grammaticalization studies thus support our hypothesis that *hinna* and *rekke* are at different stages along a grammaticalization path. As already discussed, in the verbal domain, the process where a lexical verb develops into an auxiliary is well-attested and may represent the route that both *hinna* and *rekke* are taking. As noted by Bybee (2010: 112), however, grammaticalization processes are “inherently epiphenomenal”, which means that it is difficult to sort out which processes constitute grammaticalization and which reflect common mechanisms in semantic change.

Given such difficulties, we take the different form-meaning relations found for *rekke* and view them in light of Diewald and Smirnova's (2012) 'constructionist' context-induced model for grammaticalization, in which changes along the syntagmatic axis (via specific lexico-grammatical contexts) in three distinct steps enable a fourth step in the form of the entry of a lexical item into a well-defined paradigm (in the case of *hinna* and *rekke*, the modal paradigm) on the paradigmatic axis. The main advantage with this model is that it makes it possible to distinguish grammaticalization from general semantic changes: Diewald and Smirnova (2012) argue that the first three steps are both typical of, and necessary for, grammaticalization, but not exclusive to it, while the fourth step defines grammaticalization. In our discussion of *rekke* through the lens of Diewald and Smirnova's model, we work from the assumption that synchronic data can be used to make assumptions about historical development, the idea being that diachronic change often leads to synchronic variation where a grammaticalized item may co-exist with older forms, a phenomenon known as 'layering' in grammaticalization theory (Hopper 1991: 23). Of course, such attempts at explanation remain explorative and must be complemented by diachronic work.

6.1 Steps in space-time semantic change

Diewald and Smirnova's (2012) four-step model of grammaticalization assumes a stepwise progression, which correlates with, and can be identified by, the occurrence of the relevant item (e.g. *rekke*) in particular contexts. For *rekke* (and for Swedish *hinna*, as well), we assume that the original meaning is purely spatial, representing physical, static extension, as in (16) in Section 4.2.4, reproduced here as (24).

- (24) Tora innbilte seg at dersom "huset" ikke hadde fått klippet håret
av mora, så hadde det **rukke** helt ned til hoftene. (HW1)
Tora imagined that if the house hadn't cut off her mother's hair,
it would have **reached** all the way down to her hips. (HW1T)

The first step in the stepwise semantic change of this purely spatial meaning is the occurrence of *rekke* in what Diewald and Smirnova (2012: 113) call "untypical contexts", where the item occurs in additional contexts to the original one. We believe that this context is where *rekke* occurs with a human agent (mostly) reaching out towards and possibly touching something, as in NAOB's entry (ii) (reproduced here as (25))

- (25) Han kan ikke **række** øksen (Knut Hamsun: Markens Grøde II 45, 1917)
[Lit. 'He can not reach the-axe']

In addition, we take many examples with particles like *rekke opp til* ('reach up to') to represent this extended spatial meaning, i.e. involving an agent.

The second step in the semantic change process consists of the item under change occurring in so-called “critical contexts”, rich in structural and semantic ambiguities, and thus inviting alternative interpretations (Diewald & Smirnova 2012: 113). For *rekke*, these ambiguities show up for the context above, i.e. ‘human agent + *rekke* + NP complement’, and for ‘human agent + *rekke* + adverbial complement’. Starting with the straightforward structural context where *rekke* occurs with an NP complement, as in (25) above, we note that the semantics of the complement NP determines whether *rekke* encodes a purely spatial meaning (as in 25), or the type of meaning that we have referred to as time with a spatial endpoint (e.g. in the phrase *rekke ferjen* ‘reach the ferry’, as in example (19) in Section 5, repeated here as (26)).

- (26) Han skulle **rekke** en ferje til Molde.
 [‘He was-to reach a ferry to Molde’] (LBK, SA11Mo0601.2358)

We also suggest that a number of the instances translated by *get/catch/return* are ambiguous between a spatial reading and a contextually given temporal interpretation, as in the following example from the parallel corpus.

- (27) La meg **rekke** frem i tide!” (SH1)
 Let me **get** there in time!” (SH1T)

Here, the temporal adverbial, *i tide* (‘in time’), sets up a context where *rekke* might easily be interpreted either as encoding a ‘physical path’, with a spatial endpoint defined as *frem* (‘there’ in (27)), or as a ‘temporal path’ with a temporal endpoint. In fact, the temporal interpretation is more salient. The mechanism at play here is the combination of a temporal context and *rekke*, which lends itself experientially to a temporal interpretation. This is in line with what Grady (1997) refers to as the “primary metaphor” TIME IS (MOTION ALONG) A PATH, i.e. a metaphor which has a direct experiential basis, and which motivates highly predictable sets of data. A similar context is exemplified in (28).

- (28) [...] ungdomsskoleelevane som kom med ferja frå Eikøy pustar ut
 etter springmarsjen opp til kinolokalet, det var så vidt dei **rakk** det. (EH1)
 [Lit. it was only barely they REKKE.past it]
 The junior high school students who came on the ferry from Eik
 Island are catching their breath after sprinting up to the theater
 and **getting** in just under the wire. (EH1T)

Again, a temporal understanding is close at hand. The temporal endpoint is interpreted as the beginning of the movie screening. In both example (27) and (28), we understand something beyond the mere extension of, say, an arm. Rather, the agent’s entire body moves, which presupposes an extension in time.

The third stage in Diewald and Smirnova's model is the occurrence of an item in "isolating contexts", i.e. contexts where one reading is favoured over another (Diewald & Smirnova 2012: 113). For *rekke*, we argue that the moment *rekke* occurs with an infinitival complement, a temporal reading is strongly favoured. For most speakers, it is even the only one. An example is (29).

- (29) Han ringte tre ganger før jeg *rakk* å åpne (...) (KA1)
 He rang the bell three times *before I managed* to open (...) (KA1T)

The isolating context means that *rekke* restricts the meaning to time, but a spatial reading is possible for all other contexts. Swedish *hinna*, in contrast, has gone the whole way into becoming exclusively temporal in meaning. For example, in present-day Swedish, an example like (30), corresponding to (24) above, is impossible and can only be interpreted in a context where the axe is running away and is overtaken, or where the axe is a means of transportation on a par with a train.

- (30) *Han kan inte *hinna* yxan. ['He can not HINNA the axe']

Interestingly, but outside the scope of this chapter, the Swedish verb *räcka* (obviously cognate to *rekke*) would be used in the spatial sense. If we go back in time, however, examples with a non-temporal 'reach' meaning similar to (30) are possible for *hinna*. Examples (31) and (32) are from corpora of old newspaper texts available from *Språkbanken* (Swedish Language Bank), via the search interface KÖRP.

- (31) Båten kom nu åter och kunde med möda *hinna* stranden.
 (Aftonbladet 1830)
 [Lit. 'The-boat came now back and could with effort reach the-beach.']
- (32) Nu då jag tack vare hennes upoffringar *hunnit* målet för mina långa och kostsamma studier [...] tyckte hon att min lycka var gjord
 (Aftonbladet 1840)
 [Lit. 'Now when I thanks to her sacrifices had reached the-goal for my long and expensive studies [...] thought she that my happiness was made']

Admittedly, these examples may involve more than the ability to reach something in the immediate vicinity of a human agent, but are clearly focused on the accomplishments, independent of any available time frame. Nevertheless, we speculate that these examples represent *hinna*'s occurrence in Diewald and Smirnova's (2012: 113) "critical context", involving a degree of space-time ambiguity.

An additional observation is that a majority of the cases of *rekke* with an NP complement are most naturally interpreted as denoting time sufficiency. Thus, this narrowly structural context might well also be regarded as isolating when it is embedded in a larger one that includes e.g. a temporal clause. This is especially

so when the NP itself denotes something that you would not normally physically stretch for, such as *urteforretning* ('herb store') in (33), or the more abstract *innkjøp* ('shopping') in (15), repeated here as (34).

- (33) "jeg må **rekke** en urteforretning før det stenger. Vi sees." (KA1)
 "I have to get to an herb store before closing time. See you." (KA1T)
- (34) hun **rekker** noen innkjøp før de setter seg i bilen (BV2)
 (...) she manages to do a bit of shopping before they get into the car (BV2T)

Compared to *hinna*, then, it seems that *rekke* has not completed the journey from a spatial verb into a temporal one, whereas *hinna* is exclusively temporal. The ease with which '*rekke* + NP' receives temporal interpretations indicates, however, that the Norwegian verb is moving along a similar path.

6.1.2 Paradigmatization: Is *rekke* becoming a modal?

The last stage of Diwald and Smirnova's (2012: 127f.) four-step model involves "paradigmatical (re-)integration", in our terms paradigmization, in which the grammaticalized item either is included in an already existing grammatical paradigm, or composes a new paradigm. For *rekke*, this step involves auxiliarization of *rekke* as a modal auxiliary. The prime pointer to such status is instances of *rekke* followed by a bare infinitival complement, the number one indication of modal-auxiliary status for both Faarlund et al. (1997: 526) and for Eide (2005: 56). Although our contrastive data lacks examples of '*rekke* + bare infinitive', such examples exist in monolingual corpora, even if not all Norwegian speakers accept them. In (35)–(37), we offer a few examples from NoTenTen17, Bokmål.

- (35) Brått trakk Ghevon plasmapistolen sin, og før Eena **rakk** reagere,
 hadde han avfyrt et skudd. (blogspot.no)
 [Lit. 'Suddenly drew Ghevon the-plasma-gun his, and before
 Eena REKKE.past react, had he released a shot.']
- (36) for å gjøre alt litt enklere å forstå for de som ikke har **rukket** tilegne
 seg så mye kunnskap om dette, forkorter jeg tidsaspektet til ca. 200 år
 (klimaforskning.com)
 [Lit. 'for to make everything a-little simpler to understand for those
 who not have REKKE.pp acquire so much knowledge about this, shorten
 I the-time-aspect to ca. 200 years.']
- (37) Vi må likevel starte møte så tidlig for å **rekke** informere, samt utlevere
 klær etc. til 4 lag (gti-stavanger.org)
 [Lit. 'We must anyway start meeting so early for to REKKE.inf inform,
 as-well-as deliver clothes etc to 4 teams.']

We take it that the absence of the infinitive marker in (35) to (37) indicates auxiliary use, and also note that whereas the presence of the infinitive marker is not sufficient evidence for auxiliary status, it has a clear preference for a temporal interpretation (the only possibility for many speakers).

Admittedly, *rekke* does not meet a number of other criteria for auxiliaryhood. Nevertheless, it shares some properties with Norwegian modals (which can be used both as auxiliaries and as main verbs). For example, following Eide (2005: 56ff), we find evidence in corpus data that *rekke* takes the same kinds of complements as other modals. In addition to taking a bare infinitival complement, it may occur with a past participle complement (38), a directional adverbial (39), and allows its complement to be pseudo-clefted (40) (examples from NoTenTen17, Bokmål).

- (38) tok det jeg egentlig skulle **rukket** gjort igår + litt til.
[Lit. 'took what I really was-to REKKE.pp done yesterday + some more.']
- (39) Klokka halv tre om natta må Hassan Tariq stå opp hjemme i Oslo for å **rekke** på jobb i Joker-butikken i Drammen.
[Lit. 'Clock half three in the-night must Hassan Tariq get up home in Oslo for to REKKE.inf on job in Joker-store in Drammen']
- (40) Det de har **rukket** er å starte tre forskjellige startups.
[Lit. 'What they have REKKE.pp is to start three different start-ups']

Even though especially the latter two properties demonstrate main-verb uses of Norwegian modals, our point is that *rekke* behaves like a modal in these respects. In the next section, we further motivate our assumption that *rekke* is on its way to becoming a modal (semi-) auxiliary.

6.1.3 Sufficiency and modality

In this section, we propose that in its most auxiliary-like uses, i.e. when complemented by an infinitive, *rekke* can be analyzed in terms of modality. In this use, *rekke* is like *hinna*. The latter is normally regarded as a semi-auxiliary, showing mixed behavior between main-verb and auxiliary uses. Teleman et al. call it a “potential auxiliary” (Teleman et al. 1999 (vol. 4): 297) denoting ability within a relevant space of time but lacking in a number of typical properties of auxiliaries. In our terms, *hinna* is thus a semi-auxiliary, and since ability is a modal notion, a modal semi-auxiliary. We have also argued that regardless of whether it is construed as a main verb or as an auxiliary, *hinna* expresses time sufficiency (Johansson & Nordrum 2016). Now, sufficiency, too, can be regarded as a modal notion.

Support for this can be found in analyses of sufficiency markers, which are often couched in terms of modality. For example, Meier's (2003) analysis of the

English sufficiency marker *enough* assumes a modal component to its meaning (along with a comparative one). Similarly, it is possible, as shown by Beck and Rullman (1999: 261), cited in von Fintel and Iatridou (2007), to paraphrase clauses with the adjective *sufficient* in one of two ways:

- (41) Four eggs are sufficient (to bake this cake).
- a. It is not necessary (given the rules for your cake baking) that you have more than four eggs.
 - b. It is possible (given the rules for your cake baking) that you have only four eggs

Thus, as both necessity and possibility are modal terms, it seems justified to couch sufficiency in modal terms. Johansson and Nordrum (2016) use van der Auwera and Plungian's (1998: 82) semantic map of modal meanings and view *hinna* as a participant-external sufficiency verb, i.e. as encoding possibilities enabled by some participant-external circumstance, as opposed to possibilities enabled by the internal capacities of a core participant, mostly the subject. In its auxiliary uses, then, *hinna* is a modal auxiliary; in fact, doubly so, since it includes an ability component, which would be categorized as a participant-internal component.

Turning back to *rekke*, we propose that it is on a path which *hinna* has already completed, that is, into being a modal verb with a time-sufficiency component as well as an ability component. In terms of Lehmann's (2002) parameters that characterize grammaticalization, it thus enters into a pre-existing paradigm of modal auxiliaries (i.e. the fourth step of Diewald & Smirnova's (2012) four-step model). This, in itself, does not entail either that all instances of *rekke* behave like auxiliaries (they clearly do not) nor that all instances of *rekke* have temporal meaning (they clearly do not). What it suggests instead is that there are interacting changes in progress, a semantic metaphorization process (space to time), and a structural-auxiliarization process (from adverbial/NP-complementation to VP-complementation). In the following section, we present our suggestion for how these changes are related.

6.1.4 *The grammaticalization paths of hinna and rekke*

We argue that three different semantic factors are at play in the ongoing grammaticalization of *rekke*. First, we notice a change in meaning in the form of the ubiquitous space-time metaphorization. The completion of this change is most obvious in the context of infinitival complements. It might even be argued that this context is the only structural context where *rekke* consistently and unambiguously denotes time. Other contexts like *rekke* + NP may either denote space (*rekke øksen* ('reach the-axe')) or time (*rekke toget* ('reach the-train')). In these cases, the meaning is determined by the wider discourse context, with temporal clauses, for example, resulting in temporal interpretations.

The fact that the semantic change to time is more obvious the more *rekke* behaves like an auxiliary is primarily a correlation and does not in itself entail that grammaticalization is taking/has taken place. However, we propose that temporal meanings increase the chances of grammaticalization, given the prominent role of auxiliaries in temporal reference. Moreover, as noted e.g. by Bybee and Dahl (1989), future expressions, such as *be going to* and modals expressing ‘potentiality’, are often derived from constructions with the meaning ‘movement towards a goal’. Of course, not all such expressions necessarily grammaticalize, but again, a semantic shift into the temporal domain would, we assume, facilitate grammaticalization processes. This is what we see with both *rekke* and *hinna* in the “critical contexts” where they take an NP complement. English *reach*, in contrast, has not grammaticalized, despite the fact that it is equally ambiguous in the relevant context (space in *reach the hammer*, time in *reach the store*).

The second factor we believe is at play is sufficiency. Just as for temporal expressions, expressions with source meanings involving sufficiency do not necessarily grammaticalize, but the data for *rekke* and *hinna* (as well as for Swedish *orka* ‘have enough energy to be able to do sth’), Johansson & Nordrum 2018) suggests here that where grammaticalization, specifically, auxiliarization, occurs, a likely paradigm into which a verb is attracted is that of the modal auxiliaries, which we argue above may be understood in terms of sufficiency. Moreover, van der Auwera and Plungian (1998: 94) suggest that verbs with an ‘arrive at’ meaning have the potential of entering the modal paradigm, and Bybee et al. (1994: 190) more generally list “dynamic, telic verbs” including meanings like ‘reach, arrive at’ as sources of ability/possibility modals. The latter authors (1994: 12) also make the strong claim that there is a strong relation between the meaning of the source item and the eventual grammaticalized instance, although they leave for future research the issue of the extent to which the source meaning is capable of fully predicting e.g. into what paradigm an item will grammaticalize. Our data would suggest at least that there is a strong tendency for sufficiency verbs to grammaticalize as modals.

A third possible factor is negation. In our data, the most auxiliary-like uses of *rekke*, i.e. where it takes a bare infinitive, tend to occur in negated contexts. The same observation has been made for other Norwegian verbs that take either an *å*-infinitive or a bare infinitive. Faarlund et al. (1997: 996), for example, observe about verbs like *behøve* ‘need’, *trengje* ‘need’, *orke* ‘have enough energy’, *gidde* ‘have enough volition/desire’, and *freista* ‘try’ that they occur with or without an infinitive marker in examples like those in (42)

- (42) Du behøver ikke (å) komme [‘You need NOT (to) come’]
 Vi trenger ikke (å) gå så tidlig [‘We need NOT (to) come so early’]
 Jeg orker ikke (å) høre på deg [‘I have enough energy NOT (to) listen to you’]

Han gadd ikke (å) prøve en gang
 ['He have enough desire NOT (to) try at least']
 Han freista (å) få pusten igjen ['He tried (to) get breath again']

As noted by Johannessen (1998, 2003), the optionality of the infinitive marker with the first four verbs depends on the presence of negation. In other words, taking the absence of the infinitive marker as a sign of auxiliaryhood, we might argue that negation serves as a catalyst for auxiliarization. Eide (2005: 79) includes the negated forms of *behøve* ('need') and *trengje* ('need') among Norwegian modals as modal main verbs when they take the infinitive marker and as modal auxiliaries when they occur with a bare infinitive and negation. Now, *rekke* is not as strict in requiring negation when used as an auxiliary as the verbs in (42). Nevertheless, we would speculate that, for whatever reason, negation facilitates auxiliarization.

7. Final remarks

In this chapter, we show that the most frequent English correspondences of the Norwegian verb *rekke* in the ENPC combines expressions of ability (e.g. *can*) with temporal expressions like *in time* and *before*-clauses, reflecting a complex meaning involving both ability and time. However, *rekke* is also relatively common with a concrete spatial meaning, reflected by e.g. the English verb *reach*. We further observe that, when *rekke* is complemented by an infinitive (*rekke å gjøre noe* ('have time to do sth')), the correspondences in the ENPC consistently use temporal phrases, suggesting that *rekke* in this use can only have a time-sufficiency meaning. Examples of *rekke* in monolingual corpora support this observation, and the corpora also provide examples where *rekke* is followed by a bare infinitive, thus behaving as a "true" auxiliary.

Based on our results, we suggest that *rekke* is undergoing grammaticalization from a concrete spatial verb towards one that contains a component of time sufficiency, similarly to Swedish *hinna* (Johansson & Nordrum 2016), which is further along in the grammaticalization process. We base this suggestion on the following observations: (1) Swedish *hinna* was earlier polysemous between space and time but can only denote time now; (2) Norwegian *rekke* is still polysemous between space and time, but seems to only denote time with infinitive complements.

We further assume that sufficiency in general is a modal category. As such it can be described in terms of a semantic map of modality as an instance of participant-external modality (van der Auwera & Plungian 1998). The analysis of modality in terms of participant-internal/external possibility and necessity suggests that verbs like Swedish *hinna*, and Norwegian *rekke*, which both incorporate a notion

of ability (participant-internal possibility) as well as time sufficiency (participant-external possibility) might suggest a more complex, possibly three-dimensional analysis of modality, where sufficiency is superimposed on the map of “classic” modality. We leave this observation for future research.

Acknowledgements

We thank the participants in the workshop ‘Languages in Time, Time in Languages: Phraseological Perspectives’ for useful comments and discussion, and the organizers for providing the format for a fruitful exchange of ideas. We are also indebted to the editors of this volume as well as two anonymous reviewers for many insightful comments and suggestions. We take full responsibility for all remaining inadequacies.

References

- Beck, Sigrid & Rullmann, Hotze. 1999. A flexible approach to exhaustivity in questions. *Natural Language Semantics* 7: 249–298. <https://doi.org/10.1023/A:1008373224343>
- Bhatt, Rajesh. 1999. *Covert Modality in Non-Finite Contexts*. PhD dissertation, University of Pennsylvania.
- Borin, Lars, Forsberg, Markus & Roxendal, Johan. 2012. Korp – the corpus infrastructure of Språkbanken. In *Proceedings of LREC, 2012*, 474–478. Istanbul: ELRA.
- Bybee, Joan L. 2010. *Language, Usage and Cognition*. Cambridge: CUP.
<https://doi.org/10.1017/CBO9780511750526>
- Bybee, Joan L. & Dahl, Östen. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13(1): 51–103. <https://doi.org/10.1075/sl.13.1.03byb>
- Bybee, Joan, 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.
- Chesterman, Andrew. 1998. *Contrastive Functional Analysis* [Pragmatics & Beyond New Series 47]. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.47>
- Diewald, Gabriele & Smirnova, Elena. 2012. Paradigmatic integration: The fourth stage in an expanded grammaticalization scenario. In *Grammaticalization and Language Change. New Reflections* [Studies in Language Companion Series 130], Kristin Davidse, Tine Breban, Lieselotte Brems & Tanja Mortelmans (eds), 111–133. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.130.05die>
- Eide, Kristin Melum. 2005. *Norwegian Modals* [Studies in Generative Grammar 74]. Berlin: Mouton De Gruyter.
- Faarlund, Jan Terje, Lie, Svein & Vannebo, Kjell Ivar. 1997. *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.
- Flint, Aili. 1980. *Semantic Structure in the Finnish Lexicon: Verbs of Possibility and Sufficiency*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Fortuin, Egbert. 2013. The construction of excess and sufficiency from a crosslinguistic perspective. *Linguistic Typology* 17: 31–88. <https://doi.org/10.1515/lity-2013-0002>

- Grady, Joseph. 1997. *Foundations of Meaning: Primary Metaphors and Primary Scenes*. PhD dissertation, University of California, Berkeley.
- Guevara, Emiliano Raul. 2010. NoWaC: A large web-based corpus for Norwegian. *Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop*, 1–7. Stroudsburg PA: Association for Computational Linguistics.
- Hasselgård, Hilde. 2007. Using the ENPC and the ESPC as a parallel translation corpus: Adverbs of frequency and usuality. *Nordic Journal of English Studies* 6(1). <<http://hdl.handle.net/2077/4405>> (1 April 2021). <https://doi.org/10.35360/njes.8>
- Heine, Bernd, Claudi, Ulrike & Hünemeyer, Friederike. 1991. *Grammaticalization: A Conceptual Framework*. Chicago IL: The University of Chicago Press.
- Hopper, Paul J. 1991. On some principles of grammaticization. In *Approaches to Grammaticalization, Vol. I: Focus on Theoretical and Methodological Issues* [Typological Studies in Language 19-1], Elizabeth Closs Traugott & Berndt Heine (eds), 17–35. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.19.1.04hop>
- Hopper, Paul J. & Traugott, Elizabeth Closs. 1993. *Grammaticalization*. Cambridge: CUP.
- Johannessen, Janne Bondi. 1998. Negasjonen ikke: kategori og syntaktisk posisjon. *Mons 7. Utvalde artiklar frå det 7. Møtet Om Norsk Språk i Trondheim 1997*, Jan Terje Faarlund, Brit Mæhlum & Torbjørn Nordgård (eds), 80–94. Oslo: Novus.
- Johannessen, Janne Bondi & Hagen, Kristin. 2008. *Språk i Oslo. Ny forskning omkring talespråk*. Oslo: Novus forlag.
- Johannessen, Janne Bondi. 2003. Negative polarity verbs in Norwegian. *Working Papers in Scandinavian Syntax* 71: 33–74.
- Johansson, Mats & Nordrum, Lene. 2016. Swedish *hinna* viewed through its English correspondences – Have time or be able to? *Nordic Journal of English Studies* 15: 171–199. <https://doi.org/10.35360/njes.380>
- Johansson, Mats & Nordrum, Lene. 2018. Swedish *orka* viewed through its English correspondences – Ability, insufficient strength/energy or insufficient volition? *Bergen Language and Linguistics Studies* 9(1): 155–172. <https://doi.org/10.15845/bells.v9i1.1524>
- Johansson, Stig. 2007. *Seeing through Multilingual Corpora* [Studies in Corpus Linguistics 26]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.26>
- Lehmann, Christian. 2002. *Thoughts on Grammaticalization*, 2nd, rev. edn [Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt 9]. Universität Erfurt: Christian Lehmann. <<https://www.christianlehmann.eu/publ/ASSidUE09.pdf>> (1 April 2021).
- Meier, Cécile. 2003. The meaning of *too*, *enough* and *so ... that*. *Natural Language Semantics* 11: 69–107. <https://doi.org/10.1023/A:1023002608785>
- Nadathur, Prerna. 2016. Causal necessity and sufficiency in implicativity. In *Proceedings of SALT 26 Mary Moroney*, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds), 1002–1021, <<http://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/26.1100>> (1 April 2021). <https://doi.org/10.3765/salt.v26i0.3863>
- NAOB = Det Norske Akademis ordbok. Tor Guttu (ed). Oslo: Kunnskapsforlaget og Det Norske Akademi for Språk og Litteratur. <<https://www.naob.no>> (30 November 2019).
- Nordrum, Lene. 2015. Exploring spontaneous-event marking through parallel corpora: Translating English ergative intransitive constructions into Norwegian and Swedish. *Languages in Contrast* 15(2): 230–250. <https://doi.org/10.1075/lic.15.2.04nor>
- Knudsen, Rune Lain & Vatvedt Fjeld, Ruth. 2013. LBK2013: A balanced; annotated national corpus for Norwegian Bokmål. In *Proceedings of the Workshop on Lexical Semantic Resources for NLP at NODALIDA 2013* [NEALT Proceedings Series 19, Linköping Electronic Conference Proceedings 88: 3], 12–20.

- SAOB = *Ordbok över svenska språket*. Svenska Akademien. Lund 1893–. <www.saob.se> (30 November 2019).
- Svensk-norsk blå ordbok*. 2014. Elin A. Frysjøenden (ed.). Oslo: Kunnskapsforlaget.
- Teleman, Ulf, Hellberg, Staffan & Andersson, Erik. 1999. *Svenska Akademiens Grammatik*, Vol. 1–4. Stockholm: Svenska Akademien.
- van der Auwera, Johan & Plungian, Vladimir A. 1998. Modality's semantic map. *Linguistic Typology* 1998(2): 79–124. <https://doi.org/10.1515/lity.1998.2.1.79>
- von Fintel, Kai. 2006. Modality and language. In *Encyclopedia of Philosophy*, Vol. 10, 2nd edn Donald M. Borchert (ed), 20–27. Detroit MI: MacMillan Reference.
- von Fintel, Kai & Iatridou, Sabine. 2007. Anatomy of a modal construction. *Linguistic Inquiry* 38: 445–483. <https://doi.org/10.1162/ling.2007.38.3.445>

Corpora

- Leksikografisk Bokmålskorpus (LBK)* ('Lexicographic corpus for Norwegian Bokmål'). <<https://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/lbk/index.html>> (30 May 2020).
- Norsk talespråkskorpus – Oslodelen (NoTa-Oslo)* ('Norwegian corpus of spoken language, the Oslo part'). <> (30 May 2020).
- NoTenTen17, Bokmål (Norwegian Web 2017)* available from Sketch Engine. <<https://www.sketch-engine.eu/>> (30 May 2020).
- The English-Norwegian Parallel corpus (ENPC)*. <www.hf.uio.no/ilos/english/services/omc/enpc/> (30 May 2020).
- The Norwegian Web as a Corpus v 1.0 (NoWac v 1.0)*. <<https://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/nowac/>> (30 May 2020).
- Aftonbladet* 1930, available at *Språkbanken* <<https://spraakbanken.gu.se/>> (30 May 2020).
- Aftonbladet* 1940, available at *Språkbanken*. <<https://spraakbanken.gu.se/>> (30 May 2020).

Translating verbal tenses between tensed and tenseless languages

A contrastive study of multilingual corpora

Cristina Grisot & Juan Sun

University of Geneva / Sun Yat-sen University

In this chapter, we investigate how verbal tenses are translated into a tenseless language (Chinese) and whether the translation of temporal information from a tenseless language into tensed languages (English and French) is influenced by the formal structure of the source language. The results of our corpus analysis show that English and French verbal tenses like the Present Perfect and Imparfait are not systematically translated into Mandarin using an aspectual marker. We find that present-time French and English verbal tenses are most often translated into Mandarin using non-marked VPs, past-time verbal tenses using aspectual, temporal markers and linguistically non-marked VPs, and future-time verbal tenses using temporal markers. These patterns also apply for the translation of Mandarin VPs into French and English.

1. Introduction

While the location of events and states in time is crucial to a sentence, languages are far from uniform in their means of achieving this. There are tensed languages, such as English and French, in which verbs are obligatorily inflected to express the grammatical category of ‘tense’ – a deictic category which refers to the relationship between the time of the eventuality, the time of speech and a reference time (Reichenbach 1947) – and ‘aspect’ – a grammatical category expressing how the eventuality (event or state) is viewed (Comrie 1976). This grammatical marking is usually united under the notion of ‘verbal tense’ (Comrie 1976, 1985; Dahl 1985), which refers to individual verbal tenses in particular languages (e.g. the English Simple Past or Present Perfect, or the French *Passé Composé*). Then there are tenseless, aspectual languages, such as Mandarin Chinese (henceforth,

Mandarin¹) (Li & Thomson 1981; Lin 2003, 2012; Xiao & McEnery 2004; Smith & Erbaugh 2005), in which there is no overt tense marking; here, grammatical aspect marking, like the perfective particle 了 *le* (illustrated in (1), from Smith 2008, citing Chao 1968) or the imperfective particle 着 *zhe* (illustrated in (2), from Li 2012), plays an important role in guiding the addressee to locate events and states in time.

- (1) 我摔断了腿。
Wo shuaiduan-le tui.
 [Lit. 'I break-LE leg']
 ['I broke my leg.' (and the leg is still in a cast)]
- (2) 他戴着眼镜。
Ta dai-zhe yanjing.
 [Lit. 'He wear-ZHE glasses']
 ['He is wearing glasses.']

In Mandarin, grammatical aspect is not obligatory: verbs are often used without any aspectual marker. Other mechanisms are also available, such as temporal adverbs or adverbials, and the lexical aspect of verb phrases (i.e. states and activities are unbounded and tend to be interpreted as present by default, whereas accomplishments and achievements are bounded and have default past interpretations; cf. Lin (2003) and Smith & Erbaugh (2005)). Nevertheless, for certain scholars (such as Lin (2012) and earlier publications), sentences in Mandarin are syntactically organized around an aspectual functional head, which plays the same role for locating situations in time as the tensed functional head in tensed languages.

Crucially, when translating texts between two typologically different languages (tensed and non-tensed), there is a question of how translators deal with these formal differences to preserve the semantic and the pragmatic equivalence between the source and target texts (House 2018). We expect translators to make considerable effort to preserve the content of texts while choosing the most appropriate linguistic forms in the target language to express the same content. Since both tenseless

1. The status of Mandarin as tenseless has been questioned by some scholars (Sybesma 2007; Sun 2014; Lin 2015). For Sybesma, Mandarin has a syntactic T node, albeit phonologically empty, essentially based on the premise that the presence of a T node is indispensable in determining the temporal interpretation of sentences across languages. However, such tensed analysis of Mandarin or other languages without overt tense marking “reflect a Eurocentric bias” (Shaer 2003: 139). In this study, we follow researchers such as Klein et al. (2000), Lin (2003, 2006, 2012), Smith & Erbaugh (2005), Smith (2008) and Tonhauser (2015) in treating Mandarin as a tenseless language. As we have stated previously (Sun & Grisot 2020), adopting a tenseless analysis for Mandarin allows us to assess the role of other factors in determining the temporal interpretation of sentences, such as lexical and grammatical aspect.

and tensed languages – such as Mandarin and English/French respectively – mark grammatical aspect, we then expect that grammatical aspect is a particularly relevant category for translating sentences between tensed and tenseless languages.

So far, few studies have investigated the translation of tense and aspect between tensed and tenseless languages. Using a translation corpus, Shi (2011) presents a systematic analysis of the translation of English verbal tenses into Mandarin. Shi's analysis shows that, in the corpus data analysed, all English verb phrases were translated in Mandarin using different linguistic devices (such as aspect markers, lexical verbs and context) or the same linguistic devices (such as temporal adverbs or adverbials). In particular, Shi found that English (i) perfect and progressive are frequently translated using temporal adverbs or adverbials; (ii) present-time verbal tenses are frequently translated using the context; and (iii) the Simple Past is frequently translated using aspectual markers.

This study further examines the issue by investigating the translation of verbal tenses between two tensed languages, English and French, and a tenseless language, Mandarin. Our two research questions are:

- i. How frequently are verbal tenses translated into a tenseless language using an overt aspectual marker?
- ii. Is the translation of temporal information from a tenseless language into a tensed language influenced by the formal structure of the source language?

We use translation corpora (from English and French into Mandarin, and from Mandarin into English and French) in which each of the three languages is investigated as source language (SL) and target language (TL).

This paper is structured as follows. In Section 2, we first provide an overview of the role of tense and aspect from a cross-linguistic perspective, and then briefly discuss a series of English and French perfective and imperfective verbal tenses (Sections 2.1 and 2.2); finally, we examine how time is expressed in Mandarin, with a focus on perfective and imperfective aspectual markers (Section 2.3). In Section 3, we present our empirical study, explaining our hypotheses and predictions (Section 3.1), describing the corpora and the coding system (Section 3.2), reporting the results of our quantitative analysis (3.3), and discussing them (3.4). We provide a conclusion in Section 4.

2. Tense and grammatical aspect from a cross-linguistic perspective

In tensed languages like French and English, temporal reference is expressed using verbal tenses, such as Simple Past or Present Perfect, whose meanings are provided by tight inter-relations between the categories of tense and grammatical

aspect. Tense is the “grammatical expression of location in time” (Comrie 1985: 9) and is typically expressed by inflectional morphemes. Tense is deictic as it requires reference to the speech situation (de Swart 2012). The main characteristic of tense – unlike, for example, temporal adverbials – is that it is obligatorily present in the sentence, even if temporal information is also expressed by another device, like the adverbials *yesterday* and *tomorrow*. In formal syntactic theories, the tense morpheme is assumed to occupy a functional T node, which is the head of the sentence (i.e. a string of words is not a sentence unless it contains a tensed verb). According to the ‘referential’ approach to verbal tenses (Reichenbach 1947, see also Klein 1994 from a syntactic perspective), a verbal tense expresses a relation between the moment of speech (S), the moment when the eventuality occurred (E), and a third coordinate called the reference moment R. R is an essential parameter for the temporal anchoring of situations; it can be given by the context (i.e. covert expression of R) or in the cotext by a temporal adverbial (i.e. overt expression of R), and serves to anchor eventualities temporally (Reichenbach 1947: 289).

From an aspectual perspective, the meaning of a verbal tense is described in terms of ‘grammatical aspect’, which refers to “different ways of viewing the internal temporal consistency of a situation” (Comrie 1976: 3). Grammatical aspect is not deictic and does not anchor the situation to the time axis (de Swart 2012). As typological research shows (Comrie 1976; Dahl 1985), the aspectual distinction most commonly found in languages is between perfective and imperfective aspect (Gvozdanović 2012; de Swart 2012). The perfective “indicates the view of a situation as a single whole, without distinction of the various separate phases that make up that situation” (Comrie 1976: 16); the imperfective “pays essential attention to the internal structure of the situation” (Comrie 1976: 16).

This distinction is central in Slavic languages (such as Russian and Serbian), in which every verb in the lexicon is labelled as perfective or imperfective according to the affix added to the verb stem. As such, the role of grammatical aspect in Slavic languages is pervasive. In Romance languages, tense and grammatical aspect morphemes are fused and the perfective/imperfective aspectual contrast is expressed only in the past tense. So, grammatical aspect is what constitutes the difference between the French Imparfait *il lisait* and Passé Simple *il lut* (cf. discussion in 2.2). This aspectual contrast resembles that between the English Past Progressive *he was reading* and Simple Past *he read* (cf. discussion in 2.1). Nevertheless, the English progressive expresses only part of the semantic domain of the imperfective, which is fully expressed by the French Imparfait or the imperfective in Slavic languages. Besides the perfective/imperfective aspectual contrast, languages also have a perfect/non-perfect distinction, which is viewed as a temporal operator by some scholars (Reichenbach 1947; Verkuyl 1999) and an aspectual category by others (Comrie 1976). Across languages

which have the perfect/non-perfect contrast (which excludes Slavic languages), the perfect is generally expressed periphrastically (Dahl & Velupillai 2008), as in the English Present Perfect and the French *Passé Composé*. In Mandarin, as a highly isolating aspectual language, these aspectual distinctions are expressed not by inflectional affixes on the verb, but by separate markers – aspectual particles – in the sentence (cf. 2.3).

2.1 English Simple Past, Past Progressive and Present Perfect

In English, the perfective/imperfective opposition is roughly expressed in the past time domain by the Simple Past and the Past Progressive. Classical descriptions of the English Simple Past present it as the “the deictic time preceding speech time” (Radden & Dirven 2007: 218) which has a principal temporal meaning with reference to past time (Quirk et al. 1985; Comrie 1985; Leech & Svartvik 2002; Radden & Dirven 2007). The aspectual component of the Simple Past can express both perfective and imperfective aspect, as in (3) and (4) (Huddleston & Pullum 2006). In (3), the Simple Past has a perfective interpretation, reporting a promise made in the past; in (4), it can be interpreted perfectly or imperfectly. In the former, the sentence denotes a single act of mowing the lawn located as a whole in the past time. In the latter, Sue habitually or regularly mows the lawn, and this state of affairs holds at the moment which is being referred to.

(3) I *promised* to be back for lunch.

(4) Sue *mowed* the lawn.

To investigate the rate of perfective and imperfective usages of the Simple Past, we carried out an empirical study in which we analysed the translation of the English Simple Past into Serbian (a Slavic language, in which grammatical aspect is expressed morphologically) (cf. Grisot 2017). Our hypothesis was that perfective usages of the Simple Past are translated into Serbian by perfective verbs, and imperfective usages by imperfective verbs. The analysis of 99 corpus excerpts (from literary, journalistic and parliamentary texts) containing verbs in the Simple Past showed that the Simple Past has both perfective usages (47%) and imperfective usages (53%). As such, the Simple Past does not seem to be fully perfective, and so is not in direct opposition with the Past Progressive, which represents the imperfective aspect.

Unlike the Slavic and Romance imperfectives, which express the entire imperfective semantic area (continuous and habitual readings), the English progressive does not have habitual interpretations (which are expressed by the Simple Past, as in (5)). The progressive (Past or Present) presents a strong focus on ongoing processes, restricting its application to non-stative verbs, as shown in (6) and (7) (de Swart 2012: 754).

- (5) Bill played tennis on Saturdays.
- (6) Bill was playing tennis under the rain.
- (7) *Bill was being in love with Susan.

Besides competing with the Past Progressive, the Simple Past also competes with the Present Perfect. The Present Perfect, as a compound form, expresses the perfect aspect and refers to a past time eventuality that has present time relevance. More specifically, the Present Perfect locates an eventuality in the past ($E < S$), and this is expressed from a reference moment in the present ($R = S$). The Simple Past, on the other hand, locates eventualities prior to S ($E < S$) and from a reference moment in the past ($R = E$). The main difference between the Simple Past and Present Perfect is often described as one between definite past and indefinite past (Elsness 1997). According to reference grammars, these two verbal tenses are dissimilar with respect to the respective absence and presence of a link between the past time referred to and the present time: the Simple Past implies a gap between past and present time (i.e. the two moments are disconnected), whereas the Present Perfect implies that the eventuality expressed, as in (8), continues at the present time.

- (8) The taxi has arrived.

2.2 French Passé Composé, Passé Simple and Imparfait

In French, the Passé Simple and the Imparfait are representatives of the perfective/imperfective contrast where the Passé Composé expresses the perfect aspect. In the classical approach to verbal tenses, the Passé Simple expresses a past event completed in the past with no connection to present time (Grevisse 1980: 873; Wagner & Pinchon 1962: 413). The focus on the accomplishment of the event in the past is the feature distinguishing the Passé Simple from the Passé Composé, the latter expressing a link to the speaker's or a third person's present time. The Passé Composé is described as a "tense with two faces" (Martin 1971) as it can express both past and present time: the *anteriority* Passé Composé as in (9), and the *resultative* Passé Composé as in (10).

- (9) Une fois, j'ai conduit sans le permis de conduire.
['Once, I drove without a driving licence']
- (10) Il a plu. (Saussure 2003: 232)
['It has rained.']

In a cross-linguistic typological analysis, Squartini and Bertinetto (2000) investigate the usage of the simple and compound past in Romance languages. The main hypothesis for explaining the varied usages across Romance languages is

the process of ‘aoristicization’. According to Harris (1982), the aoristicization process consists of a change from a purely perfect (such as English Present Perfect) to an aoristic form, passing through several steps: the third step corresponds to the accomplishment *Passé Composé* of the French literature, and the fourth to the anteriority *Passé Composé*. As such, the perfect aspect in Romance languages evolves diachronically towards the perfective aspect.

In its temporal usages, the *Imparfait* is classically described as a tense of background information (Weinrich 1973), aspectually unaccomplished and imperfective, which needs a hosting event previously presented (Guillemin-Flescher 1981; Kamp & Rohrer 1983; Tasmowski-De Ryck 1985; Molendijk 1990; Kleiber 2003; Berthonneau & Kleiber 1993; Vetters 1996), as shown in example (11).

- (11) Pierre entra. Marie téléphonait.
[Lit. ‘Peter entered. Mary was phoning’ (somebody)]

These features situate the *Imparfait* in opposition to the *Passé Simple*, which marks a break between the moment of speaking *S* and the global image of the eventuality happening before *S*. The *Imparfait* gives an interior perspective of the eventuality, distinguishing what has happened from what has yet to happen (Martin 1971). For other scholars, like Blücher (1974), the *Imparfait* is aspectually neutral and retrieves its aspectual value (i.e. imperfective or perfective) from the context. This was confirmed empirically by Grisot’s (2017) annotation study with corpus data, which found that the *Imparfait* was imperfective in 44.8% of cases.

2.3 Expressing time and the role of (grammatical) aspect in Mandarin

According to previous studies (Lin 2003, 2012; Smith & Erbaugh 2005; Smith 2008; Li 2012), the most important factors in the Mandarin expression of time include temporal adverbs and adverbials, overt aspectual marking (like aspectual particles and resultative verb complements), the lexical aspect of eventualities² and discourse context, among others. Below, we briefly explain the role of temporal adverbs/adverbials and lexical aspect, and then focus on overt aspectual marking (see Sun & Grisot (2020) for a more detailed description of these factors).

The reference point *R*, needed to locate eventualities in time, is most directly provided in Mandarin by temporal adverbials (Li 2012), such as various noun phrases (e.g. 本周三 *benzhousan* ‘this Wednesday’) and prepositional phrases

2. The lexical aspect of verb phrases (VPs), described in terms of Vendler’s (1957) four situation types (states, activities, achievements and accomplishments) is suggested to express default temporal reference, following the ‘Bounded Event Constraint’ principle (Smith & Erbaugh 2005; Smith 2008) or the ‘Default Aspect’ principle (Bohnenmeyer & Swift 2004).

(e.g. 于 2010 年 *yu 2010 nian* ('in 2010')). In addition, some adverbs may affect the temporal interpretation of sentences, such as past-oriented adverbs (e.g. 曾 *ceng* ('once'), 已经 *yijing* ('already')), future-oriented adverbs (e.g. 将 *jiang* ('will')), and present-oriented adverbs (e.g. 一直 *yizhi* ('always')).

Mandarin VPs are often not overtly marked for temporality or aspect. The traditional explanation for the temporal location of these VPs is the discourse context. However, Lin (2003, 2012) and Smith and Erbaugh (2005) have highlighted the role of lexical aspect in guiding addressees to locate eventualities in time, in the absence of explicit tense and aspect marking. Following Bohnemeyer and Swift's (2001, 2004) theory of 'default aspect', they propose a similar default temporal location pattern: telic or bounded eventualities (i.e. accomplishments and achievements) are located by default in the past, while atelic or unbounded ones (i.e. states and activities) are located by default in the present. As shown in examples (12) and (13), in a neutral, out-of-the-blue context, the state "be curious" is unbounded and receives default present interpretation, whereas the achievement "fall" is bounded and receives past interpretation.

- (12) 但我们仍然好奇。

Dan women renran haoqi.

[Lit. 'But we still curious']

['But we're still curious.']

- (13) 一位两岁女孩从十层住宅楼上坠落 [...]

Yiwei liangsui nühai cong shiceng zhuzhailou shang zhuiluo [...]

[Lit. 'one-CL two-year-old girl from ten-CL building up fall [...]]

['A two-year-old girl fell from a 10th floor building [...]]

To account for this default temporal pattern, Smith and Erbaugh (2005) have given a series of pragmatic principles of interpretation – the 'Deictic Principle' (the speaker is the centre of linguistic communication, and speech time S is the default orientation point; the present is located at S, the past precedes it and the future follows it), the 'Bounded Event Constraint' principle (bounded eventualities may not be located in the present, as the bounds would go beyond the present moment) and the 'Temporal Schemata Principle' (in linguistically non-marked clauses, interpret boundedness according to the temporal features of aspectual class) (see also Smith (2008) and Sun & Grisot (2020) for empirical validation).

Overt aspectual marking constrains the temporal interpretation of eventualities following the same pattern. Eventualities perfectly marked (e.g. with a perfective particle, as in (14) and (15) below) are bounded and receive past temporal interpretation, while the imperfectly marked (e.g. with an imperfective particle, as in (16) and (17)) are unbounded and receive present interpretation. The temporal interpretation arising from aspectual properties of VPs (i.e. lexical

aspect and overt aspectual marking) may be overridden by lexical items such as temporal adverbials or discourse context. The particular focus on overt aspectual marking is at the heart of the proposal that sentences in tenseless languages have in their syntactic structure an aspectual functional head, ASP, which plays the same role as the tense head in a tensed language (Lin 2012, and previous research). The aspectual head ASP-P can be perfective or imperfective, as shown in Figure 1 for a perfective ASP-P (Grisot 2018: 176).

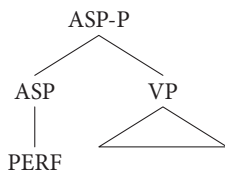


Figure 1. Syntactic structure of an aspectual functional head ASP

Overt aspect in Mandarin is commonly marked by two perfective particles (了 *le* ‘accomplished, terminative’) and 过 *guo* ‘experiential’) and two imperfective ones (着 *zhe* ‘durative’) and 在 *zai* ‘progressive’). Xiao and McEnery (2004: 91) describe 了 *le* as a perfective particle signalling “the actualization of a situation with respect to a past, present or future reference time R and present[ing] the situation as a single whole”. For Smith (2008) (following Mangione & Li 1993; Klein et al. 2000), perfective 了 *le* conveys that the event expressed by the VP occurred at some unspecified time, where perfective 过 *guo* conveys that the event occurred prior to a given time. Examples (14) and (15) (from Smith 2008, citing Chao 1968) show that these two aspectual particles code different relations between R and E. Perfective 了 *le* conveys that E=R, which means that the event is completed by the reference time (by default, S), and has an effect applying to the present. In contrast, perfective 过 *guo* conveys that E<R – that is, the event took place before R. For Smith, 过 *guo* is stative, in the sense that it focuses on the state arising once a situation has taken place. It is thus essentially a perfect, closer to the English Present Perfect or the French Passé Composé in its resultative usages (cf. Section 3.3).

- (14) 我摔断了腿。
Wo shuaiduan-le tui.
 [Lit. ‘I break-LE leg’]
 [‘I broke my leg.’ (and the leg is still in a cast)]
- (15) 我摔断过腿。
Wo shuaiduan-guo tui.
 [Lit. ‘I break-GUO leg’]
 [‘I broke my leg.’ (but the leg has since healed)]

Imperfective 在 *zai* occurs with event clauses in preverbal positions. According to Smith and Erbaugh (2005), clauses with 在 *zai* have an ongoing or progressive meaning, due to their usage with dynamic activity verbs, as in (16) (from Li 2012). The particle 着 *zhe* is more restrictive than 在 *zai*, in that it is often used with verbs of posture or movement to denote a posture or a state as a result of a dynamic activity, as in (17). Unlike 在 *zai*, clauses with 着 *zhe* are static. Both imperfective particles express that $E = R$.

- (16) 他们在打球。
Tamen zai daqiu.
 [Lit. 'They ZAI play ball']
 ['They are playing ball.']
- (17) 他戴着眼镜。
Ta dai-zhe yanjing.
 [Lit. 'He wear-ZHE glasses']
 ['He is wearing glasses.']

Aspectual information can also be conveyed by resultative verb complements (RVCs), which are formed by combining two verbs. The first one is an activity verb, and the second, as Li and Thompson (1981: 54–55) point out, “signals some result of the action or process conveyed by the first element”. As a particular construction of Mandarin, RVCs have attracted much attention from scholars (Thompson 1973; Li & Thompson 1981; Ross 1990; Li 2012). According to Li (2012), there are four main categories of RVCs: resultative RVCs (e.g. 稳 *wen* ‘stable’), achievement RVCs (e.g. 到 *dao* ‘attain’), completion RVCs (e.g. 完 *wan* ‘finish’) and directional RVCs (e.g. 走 *zou* ‘away’). RVCs are not as grammaticalized as aspectual particles, differing with respect to their large number, openness and variety of category, with grammaticalized perfective and imperfective particles instead forming a small, closed set (Li 2012). RVCs contribute to the default interpretation of eventualities because “they designate event endpoint explicitly” (Li 2012: 2053).

Furthermore, the adverbs 没/没有 *mei/ meiyou* and 未 *wei* ‘not have, not yet’ can be used to negate perfective aspect (Lü 1956). Mandarin uses the former more often than the latter, which originates from classical Chinese and occurs mostly in written Mandarin. The perfective aspect of 没/没有 *mei/ meiyou* has been discussed by many other scholars and is considered the negative form of the perfective 了 *le* (Li & Thompson 1981; Lin 2003), as shown in (18) and (19) (from Li & Thompson 1981).

- (18) 孩子撕破了那本杂志。
Haizi sipo-le neiben zazhi.
 [Lit. 'child tear-broken-LE that-CL magazine']
 ['The child tore up that magazine.']

- (19) 孩子没(有)撕破那本杂志。
Haizi mei(you) sipo neiben zazhi.
 [Lit. 'child not tear-broken that-CL magazine']
 ['The child didn't tear up that magazine.']

3. Empirical study

3.1 Hypotheses and predictions

In this study, we follow Dyvik (1998) and Noël (2003) in considering translations to be informative with respect to the linguistic meaning of source texts, despite the fact that translated texts display translation universals. Translation universals are defined as “features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems” (Baker 1993: 243). Potential translation universals include explicitation (the tendency to spell out things rather than leave them implicit in translation) and normalization (the tendency to conform to patterns and practices which are typical of the TL, or even to exaggerate them) (see also Mauranen & Kujamäki 2004). We consider that the translator's task is to preserve the semantic and pragmatic equivalence between the two languages. As such, translators choose the most appropriate formal means in the target language to render the same meaning. At the same time, if the target language displays similar formal means, they should be privileged as translation options in the target language. When this is not the case, two questions arise, as to the reason why translators chose different linguistic means in the target language, and as to what that indicates about the meaning of the expression used in the source language. As Noël (2003: 757) states, “since meanings are not directly observable [...] the translation corpus could provide a way out of this crisis, since translators, through the linguistic choices they make, inadvertently supply evidence of the meanings of the forms they are receiving and producing”. With this in mind, we make the following predictions regarding the translation of verbal tenses from English and French into Mandarin, and the choice of verbal tenses in English and French when verb phrases are translated from Mandarin into English and French.

First, we expect verb tenses with a strong aspectual component (such as the English Present Perfect, Past Progressive and Present Progressive, as well as the French Imparfait and Passé Composé) will often be translated into Mandarin with a corresponding aspectual marker: imperfective for the English Past and Present Progressive and the French Imparfait, and perfective for the English Present Perfect and the French Passé Composé. Following Noël (2003), we expect that cases in which these verbal tenses are not translated with an aspectual marker to indicate that the aspectual component of these verbal tenses is minor.

Second, we expect verb tenses with a less marked aspectual component (such as the English Simple Past and Simple Present, as well as the French *Passé Simple* and *Présent*) will often be translated into Mandarin without an aspectual marker: the Simple Past and *Passé Simple* will be translated using temporally marked VPs, and the Simple Present and the *Présent* will use linguistically non-marked VPs.

Third, we expect verb phrases translated from Mandarin into English to follow a similar pattern: aspectually marked VPs will be frequently translated using verbal tenses with a strong aspectual component, temporally marked VPs will be frequently translated using verbal tenses with a weak aspectual component, and linguistically non-marked VPs will most often be translated using present time verbal tenses, like the English Simple Present and French *Présent*.

Fourth, following the hypothesis of translation universals, we expect to find significant differences regarding the features of the original and translated texts written in each of the three languages studied. These differences should correspond to the tendency to render more explicitly and spell out messages more frequently in the translated language than in the original language (i.e. the explicitation feature), and to conform to patterns and practices which are typical of the target language, or even to exaggerate them (i.e. the normalization feature).

In summary, we expect translators to privilege the semantic and pragmatic equivalence between the two tensed languages and the tenseless language analysed here. Translators' choices will thus reflect their objective evaluation of not only the meaning communicated in the source language (verbal tenses in English and French and VPs in Mandarin) but also the most optimal linguistic means of expressing the same meaning in the target language. We are open-minded about the existence or non-existence of translation universals.

3.2 Corpora, coding system and analysis

For the corpus study discussed in this section, we built four translation corpora. The English-Mandarin corpus, shown in Table 1, consists of four English text samples (approximately 2,800 words) and their Chinese translations (approximately 5,000 words). These parallel texts were randomly collected in November 2018 from the *New York Times* online. The data were aligned at the sentence level (sentence-by-sentence). 250 VPs were then identified in English and manually extracted with their Chinese translations.

The Mandarin-English corpus consists of nine Chinese text samples (approximately 3,100 words) and their English translations (approximately 1,900 words), randomly collected in November 2018 from official websites (e.g. the Ministry of Foreign Affairs of the People's Republic of China). 200 VPs were identified in Mandarin and manually extracted with their English translations. The French-Mandarin

Table 1. General overview of the translation corpora

Corpora	Texts	Words in SL	Words in TL	VPs in SL	Sources
English-Mandarin	4	2,800	5,000	250	<i>New York Times</i> online
Mandarin-English	9	3,100	1,900	200	Website of the Ministry of Foreign Affairs of the People's Republic of China (PRC) China.org.cn
French-Mandarin	16	4,400	7,000	250	Website of the French Embassy in PRC
Mandarin-French	12	3,400	2,100	200	Website of the Chinese Embassy in France Website of Shanghai International Studies University WeChat account of CGTN French

corpus consists of 16 French text samples (approximately 4,400 words) and their Chinese translations (approximately 7,000 words), randomly collected in November 2017 from the official website of the French Embassy in China. 250 VPs were identified in French and manually extracted with their Chinese translations. The Mandarin-French corpus consists of 12 Chinese text samples (approximately 3,400 words) and their French translations (approximately 2,100 words), randomly collected in November 2017 and May 2018 from official websites (e.g. the Chinese Embassy in France). 200 VPs were identified in Mandarin and manually extracted with their French translations.

According to Smith's (2003) classification of discourse modes, most texts in the four corpora can be identified as Report; there are also passages of Narrative and Description.³ In addition, passages of quoted and reported speech were also identified. All the data were manually coded, in order to identify for each of the four directions of translation the features listed below:

3. In Smith's (2003: 26–36) classification, Narrative refers to “a sequence of events and states that have the same participants” and advances the temporal progression “with bounded events, and explicit temporal adverbials”. Report gives “an account of situations from the temporal standpoint of the reporter” and is “mainly concerned with events and states”. Passages of Description are “static and within the scope of a tacit time adverbial”.

1. The VPs in the three languages
2. The verbal tenses in French and English
3. Aspectual particles and other aspectual markers in Mandarin expressing the perfective and imperfective viewpoints
4. Linguistically non-marked VPs in Mandarin
5. The type of linguistic marking in Mandarin (temporal, aspectual or both)
6. The temporal adverbs and adverbials in Mandarin

Coded data with these features were then analysed with respect to several questions: (i) What is the distribution of French and English verbal tenses? (ii) What is the distribution of aspectually, temporally or linguistically non-marked VPs in Mandarin? (iii) What are the most frequent Mandarin translation correspondences of the French and English verbal tenses? (iv) What are the French and English translation correspondences of aspectually marked and non-marked Mandarin VPs? And finally, (v) what is the cross-linguistic picture that emerges from the answers to the previous questions?

3.3 Results

3.3.1 *Frequencies of verbal tenses in French and in English*

Table 2 provides the relative frequencies of verbal tenses in French and English source language (SL) and target language (TL). In both languages, reference to future, past and present time is most often expressed using one prototypical verbal tense.

Table 2. Frequencies of verbal tenses in French (SL and TL) and in English (SL and TL)

English verbal tenses	SL	TL	French verbal tenses	SL	TL
Simple Past	47.6%	37.7%	Présent	56%	39.7%
Simple Present	31.2%	31.2%	Passé composé	31.2%	34.8%
Present Perfect	8%	14.9%	Futur simple	7.6%	16.3%
Present Progressive	4%	0.6%	Passé simple	1.2%	1.4%
The <i>will</i> Future	2.4%	14.9%	Imparfait	2.4%	2.1%
Others	6.8%	0.6%	Others	1.6%	5.7%
Total	100%	100%	Total	100%	100%

In English SL (the 250 VPs from the English-Mandarin corpus), the two most frequent verbal tenses are the Simple Past (47.6%), expressing reference to past time, and the Simple Present (31.2%) expressing reference to present time, followed by the Present Perfect (8%) and the Present Progressive (4%). In English TL, the analysis of the Mandarin-English corpus showed that the 200 Mandarin VPs were

translated using a tensed verb phrase in 77% of cases (154 occurrences), using linguistic means other than a tensed verb phrase in 20.5% of cases (41 occurrences), and not translated (that is, zero translations) in 2.5% of cases (5 occurrences). As shown in the left-hand panel of Table 2, which reflects the translation of 154 verb phrases from Mandarin into English using a verbal tense, the distribution of verbal tenses is similar in English TL and SL. Reference to present, past and future in English TL is expressed using the Simple Past (37.7%), Simple Present (31.2%), Present Perfect (14.9%) and the *will* Future (14.9%). A Chi-Square test showed that the difference in distribution of these verbal tenses – without taking into consideration the ‘Others’ category – is not statistically significant (X^2 (4), 0.01, $p > .05$), indicating that this distribution is constant over original and translated texts.

In French SL (the 250 VPs from the French-Mandarin corpus), temporal reference to future, past and present time is expressed using one prototypical verbal tense: the Futur Simple (7.6%), Passé Composé (31.2%) and Présent (56%) respectively. In French TL, the analysis of the Mandarin-French corpus showed that the 200 Mandarin VPs were translated using a verbal tense in 70.5% of cases (141 occurrences), using linguistic means other than a tensed verb phrase in 22% of cases (44 occurrences), and not translated in 7.5% of cases (15 occurrences). As shown in the right-hand panel of Table 2, which reflects the translation of 141 VPs from Mandarin into French using a tensed verb, the distribution of verbal tenses is similar in French TL and SL. Reference to future, past and present time in French TL is expressed using the Futur Simple (16.3%), Passé Composé (34.8%) and Présent (39.7%) respectively. As in French SL, the Imparfait (2.1%) and Passé Simple (1.4%) are much less frequent than the Passé Composé. A Chi-Square test showed that the difference in distribution of these verbal tenses – without taking into consideration the ‘Others’ category – is not statistically significant (X^2 (4), 2.47, $p > .05$), indicating that this distribution is constant over original and translated texts.

3.3.2 *Frequencies of aspectual markers in Mandarin*

The analysis of our Mandarin data (SL, TL translated from English and TL translated from French) showed that aspectual markers are in fact rather infrequent compared to other temporal markers (adverbs and adverbials) and linguistically non-marked VPs (cf. Sun & Grisot (2020) for an extensive discussion on how temporal reference is expressed in Mandarin).

As shown in Figure 2, VPs marked only aspectually (the first cluster of bars) represent between 5–8% of the VPs in our data, which is constant over the three sets of Mandarin data investigated. While both aspectually and temporally marked VPs (the third cluster) are slightly more frequent than purely aspectually marked VPs, they remain much less frequent than temporally marked (second cluster) and linguistically non-marked VPs (fourth cluster).

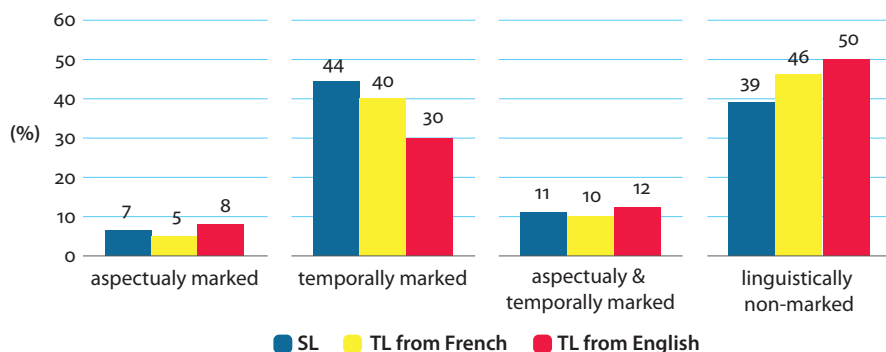


Figure 2. Frequencies of verb phrases in Mandarin SL and TL

Table 3 further elucidates the aspectual markers in our sets of Mandarin data by providing their frequencies. Among grammaticalized particles, perfective 了 *le* is by far the most frequent in Mandarin SL and TL translated from English and French. The imperfective markers 着 *zhe* and 在 *zai* are used to mark 3%-10% of the VPs. Among RVCs, the resultative 到 *dao* occurs with a rate of 11%-12% in Mandarin SL and Mandarin TL translated from French; it is much less frequent in Mandarin TL translated from English. Finally, the negative adverb 没有/未 *meiyou/wei*, having an aspectual meaning, is used more frequently in Mandarin TL than SL.

Table 3. Frequencies of aspectual markers in Mandarin (SL and TL)

		Mandarin SL	Mandarin TL (from English)	Mandarin TL (from French)
Aspectual particles	了 <i>le</i> 'accomplished, terminative'	77.4%	62%	67.6%
	着 <i>zhe</i> 'durative'	4.8%	4%	5.4%
	在 <i>zai</i> 'progressive'	-	10%	2.7%
	过 <i>guo</i> 'experiential'	3.2%	6%	-
RVCs	到 <i>dao</i> ('attain')	11.3%	2%	10.8%
	道 <i>dao</i> ('talk')	-	-	5.4%
	起 <i>qi</i> ('up')	1.6%	-	2.7%
	为 <i>wei</i> ('as')	1.6%	2%	2.7%
	至 <i>zhi</i> ('attain')	-	2%	-
	没有/未 (<i>meiyou/wei</i>) negative adverb	-	12%	2.7%
Total		100%	100%	100%

Of particular interest to the current study are the verb phrases which are aspectually marked; in our further analysis of the translation of verbal tenses into Mandarin we will consider the cases in which an aspectual marker is used alone together with those where it is used alongside a temporal marker.

3.3.3 Translation of verbal tenses from English and French into Mandarin

Figure 3 shows how the most frequent English verbal tenses from our data are translated into Mandarin. In this study, we are particularly interested in whether verbal tenses are translated using (perfective and imperfective) aspectual markers (as in Table 3). As such, the category aspectually or aspectually and temporally marked refers to VPs which are linguistically marked by an aspectual marker or by both an aspectual and a temporal marker (i.e. a temporal adverb or adverbial). The category temporally marked refers to VPs which are linguistically marked but only by a temporal marker, and thus no aspectual marker. Finally, the category linguistically non-marked refers to VPs which are marked neither by an aspectual nor by a temporal marker.

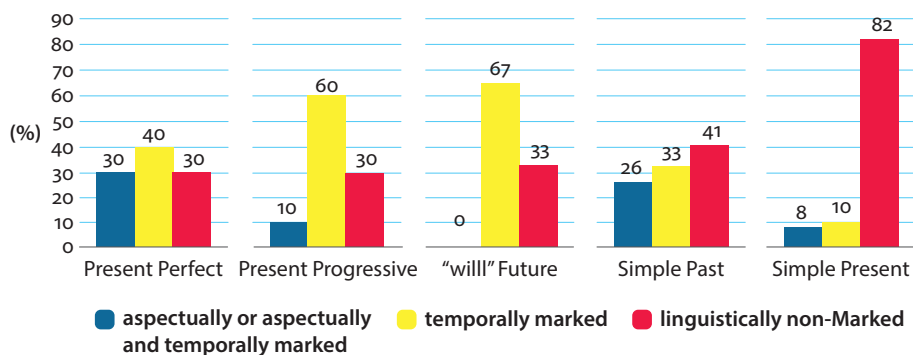


Figure 3. Translation of English verbal tenses into Mandarin

In our data, the Present Perfect and the Simple Past present a balanced distribution among the three types of VPs used in Mandarin for their translation. The Present Perfect is translated with an aspectual marker or both an aspectual and a temporal marker in 30% of cases, such as 了 *le*, 一直 *yizhi* ('always') and 着 *zhe*, as in (20), 已经 *yijing* ('already') and 了 *le*. The Present Perfect is also translated without an aspectual marker (i.e. with a temporal adverbial or adverb only) in 40% of cases, such as 一直 *yizhi* ('always'), as in (21), and with a linguistically non-marked VP in 30% of cases, as in (22).

- (20) SL: Ever since scientists created...they have braced apprehensively for [...]
 TL: 自从创造了[...]科学家们一直忧心忡忡地准备着 [...]
Zicong chuangzao-le [...] kexuejia-men yizhi youxinchongchong-de zhunbei-zhe [...]
 [Lit. 'since create-LE [...] scientist-PL always apprehensively prepare-ZHE [...]]
- (21) SL: [...]chemical plants, have been troubled by deadly leaks, fires and blasts [...]
 TL: [...]化工厂一直深受致命泄露、火灾和爆炸困扰 [...]
[...] huagongchang yizhi shenshou zhiming xielou, huozaizai he baozha kunrao [...]
 [Lit. '[...] chemical factory always deep suffer deadly leak fire and blast trouble [...]]
- (22) SL: Old cells, researchers have found, secrete proteins, lipids [...]
 TL: 研究人员发现, 衰老细胞分泌蛋白质、脂质 [...]
Yanjiuren yu xian faxian, shuailao xibao fenmi danbaizhi, Zhizhi [...]
 [Lit. 'researcher find senescence cell secrete protein lipid [...]]

Table 4 details how the most frequent English verbal tenses are translated into Mandarin.

The Simple Past is translated with an aspectual marker or with an aspectual and a temporal marker in 26% of cases, such as 过 *guo*, 已经 *yijing* ('already') and 了 *le* (as in (23)). However, the Simple Past is most frequently translated without an aspectual marker (i.e. with a temporal adverbial or adverb only) in 33% of cases, such as 当时 *dangshi* ('at that moment') (as in (24)), and with a linguistically non-marked VP in 41% of cases (as in (25)).

- (23) SL: [...] there was a woman in China carrying twins [...]
 TL: [...]中国有一位女性已经怀了双胞胎 [...]
[...] Zhongguo you yi-wei nüxing yijing huai-le shuangbaotai [...]
 [Lit. '[...] China have one-CL woman already conceive-LE twins [...]]
- (24) SL: He had the opportunity to say [...]
 TL: 当时他本有机会说 [...]
Dangshi ta ben you jihui shuo [...]
 [Lit. 'at that moment he should have opportunity say [...]]
- (25) SL: But the latest explosion was a jolting reminder [...]
 TL: 但最新的爆炸事件是一个令人震惊的提醒 [...]
Dan zuixin-de baozha shijian shi yi-ge lingrenzhenjing-de tixing [...]
 [Lit. 'But latest-DE explosion event be one-CL jolting-DE reminder [...]]

Interestingly, the Present Progressive in our data is rarely translated with an aspectual marker like 在 *zai*, as in (26). Instead translators use a temporal adverb or

Table 4. Verbal tenses in English and their translation in Mandarin

Verbal tenses in English	Markings in Mandarin	
Simple past	Aspectual particles 了 <i>le</i> ‘accomplished, terminative’	21
	过 <i>guo</i> ‘experiential’	1
	在 <i>zai</i> ‘progressive’	1
	RVCs (e.g. 到 <i>dao</i> ‘attain’)	3
	Negative adverbs 没有/未 <i>meiyou/wei</i> ‘not have, not yet’	5
	Temporal adverbs (e.g. 已经 <i>yijing</i> ‘already’)	2
	Temporal adverbials (e.g. 当时 <i>dangshi</i> ‘at that moment’)	56
	Linguistically non-marked VPs	49
	Total	119
Present perfect	Aspectual particles 了 <i>le</i> ‘accomplished, terminative’	3
	过 <i>guo</i> ‘experiential’	1
	着 <i>zhe</i> ‘durative’	1
	Negative adverb 未 <i>wei</i> ‘not have, not yet’	1
	Temporal adverbs (e.g. 已经 <i>yijing</i> ‘already’, 直 <i>yizhi</i> ‘always’)	9
	Temporal adverbials (e.g. 最近几周 <i>zuijin jizhou</i> ‘in recent weeks’)	6
	Linguistically non-marked VPs	6
	Total	20
Simple present	Aspectual particles 了 <i>le</i> ‘accomplished, terminative’	3
	在 <i>zai</i> ‘progressive’	2
	着 <i>zhe</i> ‘durative’	1
	Temporal adverbs (e.g. 通常 <i>tongchang</i> ‘generally’)	5
	Temporal adverbials (e.g. 现在 <i>xianzai</i> ‘now’)	6
	Linguistically non-marked VPs	64
	Total	78
Present progressive	Aspectual particle 在 <i>zai</i> ‘progressive’	1
	Temporal adverbs (e.g. 正 <i>zheng</i> ‘just’)	6
	Linguistically non-marked VPs	3
	Total	10
The <i>will</i> Future	Temporal adverb 将 <i>jiang</i> ‘will’	4
	Temporal adverbials (e.g. 明年 <i>mingnian</i> ‘next year’)	2
	Linguistically non-marked VPs	2
	Total	6

adverbial in 60% of cases,⁴ as in (27), or a linguistically non-marked VP in 30% of cases, as in (28).

- (26) SL: [...] investigation into the cause of the accident, are still proceeding urgently [...]
 TL: [...] 事故原因调查仍在紧张进行中 [...]
 [...] *shigu yuanyin diaocha reng zai jinzhang jinxing Zhong* [...]
 [Lit. '[...] accident cause investigation still ZAI urgently proceed middle [...]]]
- (27) SL: Dr. Nir Barzilai...is planning a study of metformin [...]
 TL: 尼尔·巴兹莱博士正计划对二甲双胍进行一项研究 [...]
Nier Bazilai boshi zheng jihua dui erjiashuanggua jinxing yi-xiang yanjiu [...]
 [Lit. 'Nir Barzilai doctor ZHENG plan DUI (Prep) metformin proceed a-CL research [...]]]
- (28) SL: Even wages are starting to rise.
 TL: 甚至工资也开始上涨。
Shenzhi gongzi ye kaishi shangzhang.
 [Lit. 'even wage also start rise']

Finally, the *will* Future and the Simple Present are primarily expressed using a temporal marker for the former, such as 将 *jiang* ('will'), as in (29), and a linguistically non-marked VP for the latter, as in (30).

- (29) SL: [...] the accident will be one of its first big tests.
 TL: [...] 这次事故将成为其首批重大考验之一。
 [...] *zhe-ci shigu jiang chengwei qi shou-pi zhongda kaoyan zhiyi.*
 [Lit. '[...] this-CL accident will become its first-CL big test one of']]
- (30) SL: And the stock markets are a mess.
 TL: 然而股票市场一团糟。
Raner gupiao shichang yituanzao.
 [Lit. 'However stock market in a mess']]

Figure 4 shows how the most frequent French verbal tenses from our data are translated into Mandarin.

4. Among the cases coded as temporally marked, it is worth noting that there are six cases where the adverbs 正 *zheng* and 正在 *zhengzai* are used to translate the Present Progressive. Despite their progressive (lexical) meaning, they are most frequently distinct from purely aspectual particles, and more similar to temporal adverbs.

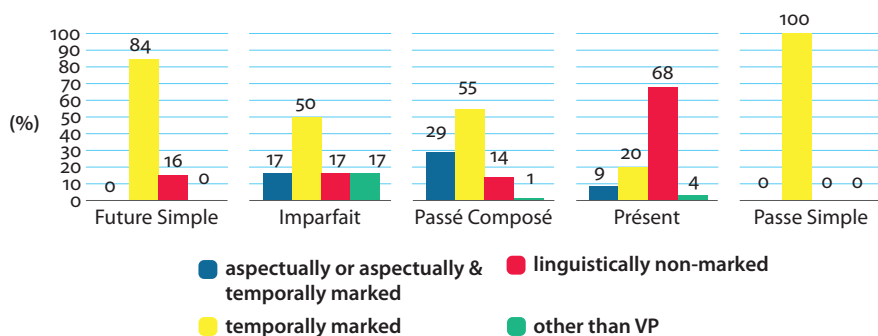


Figure 4. Translation of French verbal tenses into Mandarin

The Imparfait and the Passé Composé present a balanced distribution among the three types of VPs used in Mandarin. The Imparfait is translated using a temporal adverbial like a calendar date, as in (31), in 50% of cases. The Passé Composé is translated using a temporal marker like 过去 *guoqu* ('in the past') and 曾 *ceng* ('once'), as in (32), in 55% of cases, as well as an aspectual marker like 了 *le*, as in (33).

- (31) SL: [...] le 20 septembre 1966, atterrissait à Shanghai le premier vol direct en provenance de Paris.
 TL: [...] 1966 年 9 月 20 日, 第一班来自法国巴黎的直飞航班降落在上海机场。
 [...] 1966 nian 9 yue 20 ri, diyi-ban laizi Faguo Bali de zhifei hangban jiangluo zai shanghai jichang.
 [Lit. '[...] 1966 year 9 month 20 day first-CL from France Paris DE direct vol land in Shanghai airport']
 ['[...] on September 20, 1966, the first direct flight from Paris of France landed at Shanghai airport.']
- (32) SL: Il a exercé, en tant que magistrat, les plus hautes fonctions de l'ordre administratif français [...]
 TL: 他曾作为大法官履行法国最高行政职能 [...]
 Ta ceng zuowei dafaguan lüxing Faguo zuigao xingzheng zhineng [...]
 [Lit. 'he once as magistrate fulfil France supreme administrative function']
 ['He served as a judge to perform the highest administrative functions in France [...]]
- (33) SL: [...] la France a adopté un plan national d'action [...]
 TL: [...] 法国通过了一项国家行动计划 [...]
 [...] Faguo tongguo-le yi-xiang guojia xingdong jihua [...]
 [Lit. '[...] France adopt-LE one-CL nation action plan [...]]
 ['[...] France has adopted a national action plan [...]]

Table 5 shows how the most frequent French verbal tenses are translated into Mandarin in our set of data.

Table 5. Verbal tenses in French and their translation in Mandarin

Verbal tenses in French	Markings in Mandarin	
Passé composé	Aspectual particle 了 <i>le</i> ‘accomplished, terminative’	19
	RVCs (e.g. 到 <i>dao</i> ‘attain’)	4
	Temporal adverbs (e.g. 已经 <i>yijing</i> ‘already’, 曾 <i>ceng</i> ‘once’)	18
	Temporal adverbials (e.g. 过去 <i>guoqu</i> ‘in the past’)	52
	Linguistically non-marked VPs	11
	Other than VPs	1
	Total	78
Imparfait	Aspectual particle 了 <i>le</i> ‘accomplished, terminative’	1
	Temporal adverbials (e.g. 50 年前 <i>50 nian qian</i> ‘50 years ago’)	4
	Linguistically non-marked VPs	1
	Other than VPs	1
	Total	6
Présent	Aspectual particles 了 <i>le</i> ‘accomplished, terminative’	4
	在 <i>zai</i> ‘progressive’	1
	着 <i>zhe</i> ‘durative’	2
	RVCs (e.g. 到 <i>dao</i> ‘attain’)	4
	Negative adverb 没有 <i>meiyou</i> ‘not have, not yet’	1
	Temporal adverbs (e.g. 正 <i>zheng</i> ‘just’, 一直 <i>yizhi</i> ‘always’)	15
	Temporal adverbials (e.g. 当前 <i>dangqian</i> ‘currently’)	20
	Linguistically non-marked VPs	95
	Other than VPs	5
	Total	140
Futur simple	Temporal adverb 将 <i>jiang</i> ‘will’	15
	Temporal adverbials (e.g. 届时 <i>jieshi</i> ‘at that moment’)	10
	Linguistically non-marked VPs	3
	Total	19

The Présent is most frequently translated using a linguistically non-marked VP (68%), as in (34). In 20% of cases, a temporal marker is used, such as 一直 *yizhi* ‘always’, as in (35).

- (34) SL: Mais ce n'est pas suffisant [...].
 TL: 但这还是不够的 [...]
Dan zhe haishi bugou-de [...]
 [Lit. ‘But this still be not enough-DE [...]]’
 [‘But this is not enough [...]]’

- (35) SL: [...] la France tient à rester le partenaire privilégié de la Chine [...]
 TL: [...] 法国 [...] 一直是中国的特殊伙伴。
 [...] *Faguo* [...] *yizhi shi Zhongguo-de teshu huoban* [...]
 [Lit. ' [...] France [...] always be China-DE special partner [...]]
 [' [...] France remains the privileged partner of China [...]]

The French Futur Simple and Passé Simple are most often translated using a temporal marker like 将 *jiang* ('will'), as in (36), for the former and 曾 *ceng* ('once'), as in (37), for the latter.

- (36) SL: Elle entrera en fonction le 15 novembre.
 TL: 她将于11月15日就职。
Ta jiang yu 11 yue 15 ri jiuzhi.
 [Lit. 'she will YU (prep) 11 month 15 day take office']
 ['She will take office on November 15']
- (37) SL: Il fut enseignant dans plusieurs établissements français d'enseignement supérieur [...]
 TL: 他曾在法国多所高等教育机构任教 [...]
Ta ceng zai Faguo duo-suo gaodeng jiaoyu jigou renjiao [...]
 [Lit. 'he once in France some-CL higher education institution teach [...]]
 ['He was a teacher in several French higher education institutions [...]]

3.3.4 Translation of verb phrases from Mandarin into English and French

The analysis of the Mandarin source texts in the Mandarin-English corpus shows that 38% of the 200 VPs investigated are linguistically non-marked. The rest are aspectually or aspectually and temporally marked in 14% of cases, and temporally marked in 48% of cases. Figure 5 indicates that Mandarin VPs in our set of data which are temporally marked with an adverb like 已经 *yijing* ('already') are translated using the Present Perfect (14%) or Simple Past (38%), whereas temporally marked VPs with an adverb like 正 *zheng* ('just') are translated using the Simple Present (20%).

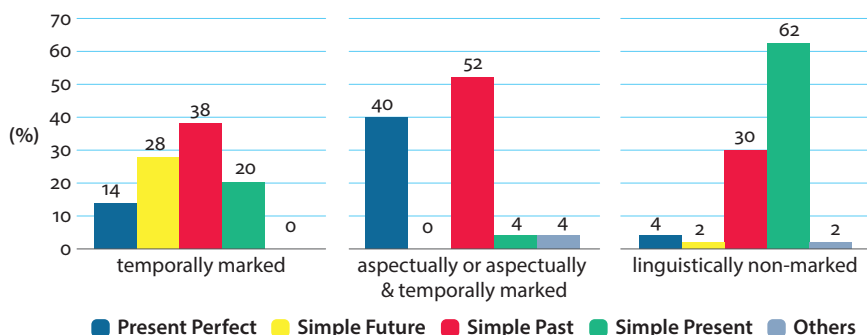


Figure 5. Translation of Mandarin VPs into English

VPs marked with an aspectual marker alone or both an aspectual and a temporal marker are most frequently translated using the Simple Past (52%) or Present Perfect (40%). For example, as Table 6 shows, VPs marked by the perfective particle 了 *le* are translated using the Simple Past (12 occurrences) or the Present Perfect (7 occurrences) (especially those marked with both 已经 *yijing* ‘already’ and 了 *le*). The linguistically non-marked VPs in our data are translated most frequently with the Simple Present (62%), followed by the Simple Past (30%). The Present Perfect and the “will” Futures are much less frequently used.

Table 6. Aspectual markers in Mandarin and verbal tenses in English

Aspectual markers in Mandarin		Verbal tenses in English	
Perfective	Aspectual particle 了 <i>le</i> ‘accomplished, terminative’	Simple past	12
		Present perfect	7
		Other than VPs	3
	Aspectual particle 过 <i>guo</i> ‘experiential’ RVCs 到 <i>dao</i> ‘attain’ 为 <i>wei</i> ‘as’	Present perfect	1
		Present perfect	2
		Simple past	1
Imperfective	Aspectual particle 着 <i>zhe</i> ‘durative’	would Future	1
	Total	Simple present	1
			28

The analysis of the Mandarin source texts in the Mandarin-French corpus shows that, of the 200 VPs investigated, 39% are linguistically non-marked, 44% are aspectually non-marked, and 17% are aspectually or aspectually and temporally marked. Unlike the Mandarin-English translation direction, temporally marked VPs present a balanced distribution of the verbal tenses used in French. VPs marked with the adverb 将 *jiang* ‘will’ are translated using the Futur Simple, VPs marked with an adverbial such as 日前 *riqian* ‘recently’ are translated using the Passé Composé, and 现在 *xianzai* ‘now’ or 正在 *zhengzai* ‘in process of’ are translated using the Présent (Figure 6).

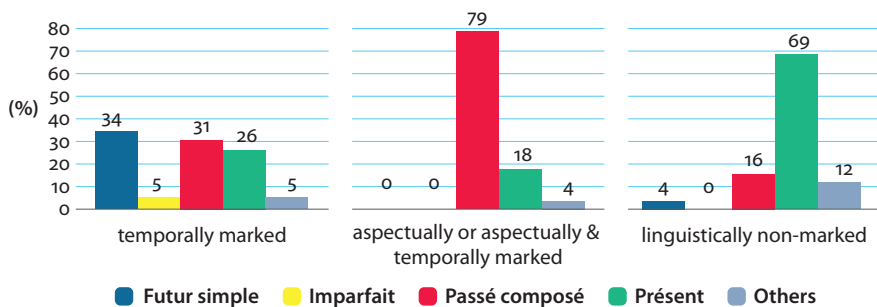


Figure 6. Translation of Mandarin VPs into French

VPs marked with an aspectual marker alone or both an aspectual and a temporal marker are most frequently translated by the *Passé Composé* (79% of cases) or the *Présent* (18% of cases). As Table 7 shows, the *Passé Composé* is used to translate VPs marked by aspectual markers such as 了 *le*, 过 *guo* and 到 *dao* ('attain'), whereas the *Présent* is used in the case of aspectual markers like 起 *qi* ('up') and 着 *zhe*. Finally, non-linguistically marked VPs are translated using the *Présent* (69%) or the *Passé Composé* (16%).

Table 7. Aspectual markers in Mandarin and verbal tenses in French

Aspectual markers in Mandarin		Verbal tenses in French	
Perfective	Aspectual particle 了 <i>le</i> 'accomplished, terminative'	<i>Passé composé</i>	17
		<i>Présent</i>	2
		Plus-que-parfait	1
		Other than VPs	6
	Aspectual particle 过 <i>guo</i> 'experiential'	<i>Passé composé</i>	1
	RVCs 到 <i>dao</i> ('attain')	<i>Passé composé</i>	4
Imperfective	起 <i>qi</i> ('up')	<i>Présent</i>	1
	Aspectual particle 着 <i>zhe</i> 'durative'	<i>Présent</i>	2
	Total		34

3.4 Discussion of the results

3.4.1 *Frequencies of verbal tenses in English and French*

The results of our quantitative analysis of the four translation corpora investigated reveal that the most frequent English verbal tenses in our sets of data are the Simple Past, the Simple Present and the Present Perfect. This distribution is similar in English SL and TL, although the *will* Future is more frequent in English TL than in English SL (14.9% vs. 2.4%). As noted below, we attribute this to the translation into English of Mandarin VPs overtly marked with the future-oriented adverb 将 *jiang* ('will') – frequent in Mandarin SL – using the *will* Future.

In the French sets of data, we found that the most frequent verbal tenses are three prototypical tenses expressing present, past and future time: the *Présent*, the *Passé Composé* (in competition with the *Passé Simple*, the latter being rather infrequent) and the *Futur Simple*. This distribution was found to be similar in French SL and TL, indicating that there is no strong influence of the formal structure of the source language (Mandarin) on the target language. In both French SL and TL, the frequencies of the *Imparfait* and *Passé Simple*, which can also be described as E<S, are much lower than the prototypical past tense, the *Passé Composé*. The data from French SL and TL illustrate the aoristicization process undergone by

the French Passé Composé, which is the most frequently used perfective past time verbal tense. This has led to the decline of the Passé Simple, often observed in contemporary French in several stylistic registers, with the notable exception of literary texts (as confirmed by Grisot's (2018) corpus study).

3.4.2 *Frequencies of aspectual markers in Mandarin*

The analysis of the Mandarin SL and TL texts shows that the most frequent type of verb phrases in our data are linguistically non-marked VPs and temporally marked VPs. Aspectually marked VPs are much less frequent, as are VPs marked by both an aspectual and a temporal marker. This result is surprising given the proposal that Mandarin is an aspectual language, which has in its syntactic structure an aspectual functional head ASP playing the same role as the tense head in a tensed language like English and French (Lin 2012, and previous research). If this were true, we would expect to find VPs marked with an overt (perfective or imperfective) marker particularly frequently. Instead, our results show that temporal reference is most frequently expressed in Mandarin using a temporal adverb and adverbial starting from a default localization in the present time of linguistically non-marked VPs following the Deictic Principle (Smith 2008). Indeed, Sun and Grisot (2020) found that, unlike temporal markers and the Deictic Principle applying to linguistically non-marked VPs, aspectual markers do not play a statistically significant role in the expression of temporal reference in Mandarin.

Furthermore, this distribution of temporally, aspectually and non-linguistically marked VPs is similar in Mandarin SL, TL translated from English and TL from French, indicating that these translations do not present the alleged universal features of explicitation or normalization. If this were the case, we would have found significantly more aspectually marked VPs in Mandarin TL (translated from English and from French) than in Mandarin SL.

3.4.3 *Translation of English and French verbal tenses into Mandarin*

Contrary to expectations, the Present Perfect – which has a strong aspectual component – is translated in our data using an aspectually marked VP in only 30% of cases. Here, the most frequent aspectual marker is 了 *le*, which has an accomplished or terminative aspectual meaning. In our corpus, 过 *guo* with an experiential aspectual meaning is used only once to translate the Present Perfect. Often, these aspectual markers do not occur alone but mark a VP at the same time with a temporal adverb, such as 已经 *yijing* ('already') or 一直 *yizhi* ('always'), or a temporal adverbial, such as 最近几周 *zuijin jizhou* ('in recent weeks'). Yet the Present Perfect was also translated using only a temporal adverbial or a linguistically non-marked VP. We suggest that these three types of translation inform us about the semantics and pragmatics of the Present Perfect. On the basis of our set of data,

the Present Perfect expresses with equal frequency the perfect (accomplished or terminative) aspect (as shown by its translation with aspectually marked VPs) and reference to past and to present time (as shown by its translation by temporally marked and linguistically non-marked VPs).⁵

The Simple Past presents a similar case: it is translated into Mandarin using aspectually marked VPs, temporally marked VPs and linguistically non-marked VPs with equal frequency. This array of translation choice illustrates that the meaning of the Simple Past is less definite than classical approaches predict. As Section 2 states, the Simple Past is described as representing an action or state as having occurred or having existed at a past moment or during a past period of time that is definitely separated from the actual present moment of speaking or writing. In our data, the Simple Past is translated using a linguistically non-marked VP in 41% of cases. Knowing that linguistically non-marked VPs most frequently express present time (Sun & Grisot 2020), we may now question the fact that the Simple Past expresses past eventualities completely detached from the present time. As for its aspectual component, its translation using the particle 了 *le* shows that the Simple Past and the Present Perfect share these accomplished or terminative aspectual types of meaning. This finding is in line with the ‘competition’ thesis between these two English verbal tenses. Unexpectedly, the Simple Past is only very rarely translated using an imperfective aspectual particle, the progressive 在 *zai*, which points to the rather reduced imperfective meaning of the Simple Past.

The picture is quite clear in the case of Simple Present and the *will* Future. The former is translated in 82% of cases using a linguistically non-marked VP. Its semantics as E=R=S is perfectly paralleled by a translation with a linguistically non-marked VP, whose temporal reference is given by the Deictic Principle. The latter tense is most frequently translated with a VP marked by the temporal adverb 将 *jiang*, which roughly means ‘will’.

In the French set of data, we found that the *Passé Composé* is most frequently translated using temporal adverbs, such as 已经 *yijing* (‘already’) and 曾 *ceng* (‘once’), and temporal adverbials, such as 过去 *guoqu* (‘in the past’). This type of translation illustrates the fact that the meaning of this verbal tense is mainly oriented around its past (E<S) temporal component. Furthermore, the translation

5. As Sun and Grisot (2020) find, linguistically non-marked VPs most frequently express reference to present time, followed in decreasing order by reference to past time, and to future time. This temporal information is inferred on the basis of the lexical aspect of the VP: events are most frequently located in the past, activities are located in the past or the future with equal frequency, and states are most frequently located in the present. We found these patterns in Mandarin SL, in Mandarin TL translated from English, and in Mandarin TL translated from French.

using the accomplished or terminative 了 *le* confirms the aspectual component of the Passé Composé: it expresses accomplished or terminated actions. Experiential 过 *guo* is never used in our data to translate the Passé Composé, which is surprising given that this verbal tense is frequently used to express experiences (as in *J'ai gravi l'Everest* ('I climbed Everest')).

As in the English data, the Présent and the Futur Simple do not present unexpected translation patterns. The former is most frequently translated using a linguistically non-marked VP or a VP temporally marked with temporal adverbs, such as 正 *zheng* ('just') and 一直 *yizhi* ('always'), and adverbials such as 当前 *dangqian* ('currently'). The latter is most frequently translated using the temporal adverb 将 *jiang* ('will').

In contrast, the Imparfait's translation is surprising. Given that most linguistic studies describe it as an imperfective verbal tense, we would expect it to be most often translated using an imperfective aspectual marker – yet in our corpus, none of the rare occurrences of the Imparfait were translated in this way, and its translation possibilities in fact highlight its past time temporal meaning. This finding is in line with the results of Grisot's (2018) annotation study, where the Imparfait was annotated as imperfective in under 50% of cases.

3.4.4 Translation of Mandarin verb phrases into English and French

In general, the translation of Mandarin verb phrases into English and French confirms the above-mentioned findings regarding the meaning of English and French verbal tenses. In particular, we found that aspectually or aspectually and temporally marked VPs in the Mandarin-English corpus were most frequently translated using the Present Perfect or the Simple Past, indicating that these two verbal tenses have a rather solid aspectual component compared to other verbal tenses. Looking at the aspectual markers themselves, we find that the particle 了 *le* may be translated either by the Simple Past or the Present Perfect.

In contrast, only temporally marked VPs were translated using a larger array of verbal tenses, such as the Present Perfect, the *will* Future, the Simple Past and the Simple Present. Finally, the 'linguistically non-marked VPs – Simple Present' pattern found in the English-Mandarin corpus was confirmed for the Mandarin-English direction of translation.

The comparison with French in the Mandarin-French corpus shows very similar results. We found that aspectually or aspectually and temporally marked VPs were most often translated using the Passé Composé and the Présent. Again, these results correspond to the translation of the particle 了 *le*, which is most often translated using a Passé Composé. Furthermore, temporally marked VPs were translated using several verbal tenses (Futur Simple, Passé Composé, Présent and

Imparfait) whereas linguistically non-marked VPs were translated by means of the *Présent* or the *Passé Composé*.

We found no significant differences between the source and target versions of each language, with the exception of the *will* Future (cf. Table 2). We think that the higher frequency of the *will* Future in English TL compared to English SL is corpus-dependent, tightly linked to a higher frequency of the future-oriented adverb 将 *jiang* in the Mandarin-English corpus; otherwise, we would have found the same pattern in the Mandarin-French corpus.

4. Conclusion

In this study, we have illustrated how translation corpora may be used to investigate the meaning of English and French verbal tenses via their translation into a tenseless language such as Mandarin, and also via the translation of Mandarin VPs into English and French. Our results can be summarized as follows.

First, our quantitative analyses show that the translation of verbal tenses follows patterns which are constant across the languages analysed and the translation directions analysed. For example, in harmony with Shi's (2011) English-Mandarin corpus investigation, we found that the perfect aspect (i.e. Present Perfect) is most often translated using temporal adverbs and adverbials, and that present-time verbal tenses (i.e. Simple Present) are most often translated using linguistically non-marked VPs. In terms of past-time verbal tenses (i.e. Simple Past), translated most frequently in Shi's corpus using aspectual markers, the Simple Past is translated using temporally marked, aspectually marked and linguistically non-marked VPs with equal frequency in our set of data. Future-time verbal tenses are most frequently translated using temporal markers, such as the adverb 将 *jiang*. These patterns were confirmed in both the French-Mandarin corpus and in the translation of Mandarin VPs into English and French.

Second, aspectual markers in our sets of data were much less frequent than temporal markers and linguistically non-marked VPs. This result does not confirm the expectation drawn from Lin's (2012) hypothesis of the functional aspectual head in Mandarin, according to which grammatical aspect is expected to be an essential category when translating between tensed and tenseless languages, and translators should privilege aspectual markers when translating verbal tenses with a strong aspectual component. Yet our data do not support this hypothesis, either in the case of English verbal tenses such as the Present Perfect and Present Progressive or French verbal tenses such as the *Imparfait*. As already pointed out in the literature (cf. de Swart 2012), aspect is a complex category which varies from

one language/family of languages to another. For a more comprehensive picture of the status of aspect, from both formal and semantic points of view, further systematic comparisons using translation corpus data should be undertaken between Romance, Germanic, Slavic and Sino-Tibetan languages.

Third, we did not find evidence supporting translation universals in the case of the translation of tense and aspect, such as explicitation and normalization features. According to this hypothesis, we should have observed in the translated texts that the original message is spelled out, and that the formal features of the target language are especially present, or even exaggerated; but we found no significant differences with respect to the distribution and translation of verbal tenses. We suggest that future research should be carried out to investigate further the unexpected pattern found for the translation of aspectual verbal tenses, such as the Present Progressive, the Past Progressive and the Imparfait. Due to their reduced frequency in our sets of data, our findings need to be confirmed in a larger corpus.

Acknowledgements

This research has been carried out within the Youth Program of Humanities and Social Sciences Foundation of the Ministry of Education of the People's Republic of China (No. 18YJCZH155), as well as the Swiss National Science Foundation for the VTS Project ("Verbal tenses and subjectivity: an empirical cognitive approach; No. 100015_170008/1").

References

- Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In *Text and Technology: In Honour of John Sinclair*, Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds), 233–253. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.64.15bak>
- Berthonneau, Anne-Marie & Kleiber, Georges. 1993. Pour une nouvelle approche de l'imparfait: L'imparfait, un temps anaphorique méronomique. *Langages* 112: 55–73. <https://doi.org/10.3406/lgge.1993.1661>
- Blücher, Kolbjørn. 1974. L'imperfetto italiano è veramente un imperfetto? In *Actes du XIIIe Congrès International de Linguistique et Philologie Romanes*, 397–402. Québec: Les Presses de l'Université Laval.
- Bohnenmeyer, Jürgen & Swift, Mary. 2001. Default aspect: The semantic interaction of aspectual viewpoint and telicity. In *Proceedings from the Perspectives on Aspect*. Utrecht: Institute of Linguistics.
- Bohnenmeyer, Jürgen & Swift, Mary. 2004. Event realization and default aspect. *Linguistics and Philosophy* 27(3): 263–296. <https://doi.org/10.1023/B:LING.0000023371.15460.43>
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley CA: University of California Press.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: CUP.

- Comrie, Bernard. 1985. *Tense*. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139165815>
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
- Dahl, Östen & Velupillai, Viveka. 2008. Perfective/imperfective aspect. In *World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds). Munich: Max Planck Digital Library.
- Dyvik, Helge. 1998. A translational basis for semantics. In *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, Stig Johansson & Signe Oksefjell (eds), 51–86. Amsterdam: Rodopi.
- Elsness, Johan. 1997. *The Perfect and the Preterite in Contemporary and Earlier English*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110810264>
- Grevisse, Maurice. 1980. *Le bon usage*, 11^e édn. Bruxelles: De Boeck Duculot.
- Grisot, Cristina. 2017. Tense, grammatical aspect and subjectivity: An experimental study using inter-annotator agreement rates and corpus-based data. *Corpus Pragmatics* 2(1): 27–55. <https://doi.org/10.1007/s41701-017-0021-z>
- Grisot, Cristina. 2018. *Cohesion, Coherence and Time from an Experimental Corpus Pragmatics Perspective*. Cham: Springer.
- Guillemín-Flescher, Jacqueline. 1981. *Syntaxe comparée du français et de l'anglais: problèmes de traduction*. Paris: Éditions Ophrys.
- Gvozdanović, Jadranka. 2012. Verbal aspect. In *The Oxford Handbook of Tense and Aspect*, Robert Binnick (ed.), 752–780. Oxford: OUP.
- Harris, Martin (1982). The “past simple” and “present perfect” in Romance. In *Studies in the Romance Verb*, Martin Harris & Nigel Vincent (eds), 42–70. London: Croom Helm.
- House, Juliane. 2018. *Translation: The Basics*. London: Routledge.
- Huddleston, Rodney D. & Pullum, Geoffrey K. 2006. *A Student's Introduction to English Grammar*, 2nd edn. Cambridge: CUP.
- 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: Walter de Gruyter. <https://doi.org/10.1515/9783110852820.250>
- Kleiber, Georges. 2003. Entre les deux mon cœur balance ou l'imparfait entre aspect et anaphore. *Langue Française* 138: 8–19. <https://doi.org/10.3406/lfr.2003.6478>
- Klein, Wolfgang. 1994. *Time in Language*. London: Routledge.
- Klein, Wolfgang, Li, Ping & Hendriks, Henriette. 2000. Aspect and assertion in Mandarin Chinese. *Natural Language & Linguistic Theory* 18(4): 723–770. <https://doi.org/10.1023/A:1006411825993>
- Leech, Geoffrey & Svartvik, Jan. 2002. *A Communicative Grammar of English*, 3rd edn. New York NY: Routledge.
- Li, Charles N. & Thompson, Sandra A. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley CA: University of California Press.
- Li, Wendan. 2012. Temporal and aspectual references in Mandarin Chinese. *Journal of Pragmatics* 44(14): 2045–2066. <https://doi.org/10.1016/j.pragma.2012.09.017>
- Lin, Jo-Wang. 2003. Temporal reference in Mandarin Chinese. *Journal of East Asian Linguistics* 12(3): 259–311. <https://doi.org/10.1023/A:1023665301095>
- Lin, Jo-Wang. 2006. Time in a language without tense: The case of Chinese. *Journal of Semantics* 23(1): 1–53. <https://doi.org/10.1093/jos/ffh033>
- Lin, Jo-Wang. 2012. Tenselessness. In *The Oxford Handbook of Tense and Aspect*, Robert I. Binnick (ed.), 669–695. Oxford: OUP.

- Lin, Tzong-Hong Jonah. 2015. Tense in Mandarin Chinese sentences. *Syntax* 18(3): 320–342. <https://doi.org/10.1111/synt.12032>
- Lü, Shuxiang. 1956. *Zhongguo wenfa yaolue* (A Concise Chinese Grammar). Beijing: The Commercial Press.
- Mangione, Louis & Dingzuan, Li. 1993. A compositional analysis of - *guo* and - *le*. *Journal of Chinese Linguistics* 21(1): 65–122.
- Martin, Robert. 1971. *Temps et aspect. Essai sur l'emploi des temps narratifs en moyen français*. Paris: Klincksieck.
- Mauranen, Anna & Kujamäki, Pekka. 2004. *Translation Universals: Do They Exist?* [Benjamins Translation Library 48]. Amsterdam: John Benjamins. <https://doi.org/10.1075/btl.48>
- Molendijk, Arie L. 1990. *Le passé simple et l'imparfait: Une approche reichenbachienne*. Amsterdam: Rodopi.
- Noël, Dirk. 2003. Translations as evidence for semantics: An illustration. *Linguistics* 41(4): 757–785. <https://doi.org/10.1515/ling.2003.024>
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Radden, Günter & Dirven, René. 2007. *Cognitive English Grammar*. Amsterdam: John Benjamins. <https://doi.org/10.1075/clip.2>
- Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York NY: McMillan.
- Ross, Claudia. 1990. Resultative verb compounds. *Journal of the Chinese Language Teachers Association* 25(3): 61–83.
- de Saussure, Louis. 2003. *Temps et pertinence: Éléments de pragmatique cognitive du temps*. Bruxelles: De Boeck Duculot. <https://doi.org/10.3917/dbu.sauss.2003.01>
- Shaer, Benjamin. 2003. Toward the tenseless analysis of a tenseless language. In *Proceedings of the 2nd Conference on the Semantics of Under-Represented Languages in the Americas [SULA 2]*, Jan Andersson, Paula Menéndez-Benito & Adam Werle (eds), 139–156. Amherst MA: GLSA.
- Shi, Mimi. 2011. A Corpus-based Contrastive Approach for the Analysis of Tense and Aspect in Translation from English into Mandarin Chinese. PhD dissertation, Westfälische Wilhelms-Universität Münster.
- Smith, Carlota S. 2003. *Modes of Discourse: The Local Structure of Texts*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511615108>
- Smith, Carlota S. 2008. Time with and without tense. In *Time and Modality*, Jacqueline Guéron & Jacqueline Lecarme (eds), 227–249. Berlin: Springer. https://doi.org/10.1007/978-1-4020-8354-9_10
- Smith, Carlota S. & Erbaugh, Mary S. 2005. Temporal interpretation in Mandarin Chinese. *Linguistics* 43(4): 713–756. <https://doi.org/10.1515/ling.2005.43.4.713>
- 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.), 403–440. Berlin: Mouton de Gruyter.
- Sun, Hongyuan. 2014. Temporal Construals of Bare Predicates in Mandarin Chinese. PhD dissertation, Leiden University.
- Sun, Juan & Grisot, Cristina. 2020. Expressing temporal reference in Mandarin: A quantitative study using translation corpora. *Corpus Pragmatics* 4: 207–241. <https://doi.org/10.1007/s41701-020-00077-x>
- de Swart, Henriette. 2012. Verbal aspect. In *The Oxford Handbook of Tense and Aspect*, Robert Binnck (ed.), 752–780. Oxford: OUP.

- Sybesma, Rint. 2007. Whether we tense-agree overtly or not. *Linguistic Inquiry* 38(3): 580–587.
<https://doi.org/10.1162/ling.2007.38.3.580>
- Tasmowski-De Ryck, Liliane. 1985. L'imparfait avec et sans rupture. *Langue Française* 67: 59–77.
<https://doi.org/10.3406/lfr.1985.4651>
- Thompson, Sandra A. 1973. Resultative verb compounds in Mandarin Chinese: A case for lexical rules. *Language* 49(2): 361–379. <https://doi.org/10.2307/412459>
- Tonhauser, Judith. 2015. Cross-linguistic temporal reference. *Annual Review of Linguistics* 1(1): 129–154. <https://doi.org/10.1146/annurev-linguist-030514-124923>
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2): 143–160.
<https://doi.org/10.2307/2182371>
- Verkuyl, Henk. 1999. *Aspectual Issues: Studies on Time and Quantity*. Stanford CA: CSLI.
- Vetters, Carl. 1996. *Temps, Aspect et Narration*. Amsterdam: Rodopi.
- Wagner, Robert Léon & Pinchon, Jacqueline. 1962. *Grammaire du français classique et moderne*. Paris: Hachette.
- Weinrich, Harald. 1973. *Le temps*. Paris: Seuil.
- Xiao, Richard & McEnery, Tony. 2004. *Aspect in Mandarin Chinese: A Corpus-based Study* [Studies in Language Companion Series 73]. Amsterdam: John Benjamins.
<https://doi.org/10.1075/slcs.73>

From language to language, from time to time

Echoic binomials from an English-German-Swedish perspective

Magnus Levin & Jenny Ström Herold

Linnaeus University

This study is situated within the broader field of phraseology and concerns repetitive, echoic binomials such as *day by day* (NPN) and *on and on* ('ADV and ADV'). While the bulk of previous research has focused on their use in individual languages (Jackendoff 2008; Ziem 2008), this study takes a comparative approach, using data from two parallel corpora: the Linnaeus University English-German-Swedish corpus and the English-Swedish Parallel Corpus. The results indicate that binomials, in particular of the NPN type, are most frequent in Swedish originals. In English originals, they are rare and not very productive. In translations, frequencies closely mirror those in their originals, but other recurrent patterns (e.g., *one X at a time*) are also commonly used as correspondences.

1. Introduction

Echoic, or repetitive binomials (Mollin 2013: 172, Gustafsson 1975: 9) are a cross-linguistic phenomenon dating back at least to biblical times (e.g., *tooth for tooth*) (Malkiel 1959: 125–126). Such binomials are characterized by a tripartite 'XYX' structure in which two identical lexical items are linked by the coordinator *and* or a preposition. The repeated items may be nouns, as in the NPN type *day by day* discussed by Jackendoff (2008), adverbs as in *again and again* (cf. Levin & Lindquist 2013), adjectives (*smaller and smaller*) or verbs (*waited and waited*). Although such constructions are attested cross-linguistically (Malkiel 1959; Jackendoff 2008: 8), contrastive studies are lacking.

Echoic binomials express a variety of meanings, but most have a temporal or dynamic interpretation (Knappe 2016), even though the repeated items in themselves do not always denote temporal concepts. In (1), motion across space entails motion in time, and (2) refers to repeated events in time.

- (1) [...] in the seventeenth century, the horse found a new role moving ordnance quickly *from place to place* during battles. (LEGS; EN original)
 [...] im 17. Jahrhundert fanden Pferde eine neue Aufgabe, indem sie die Geschütze während der Schlachten schnell *von Stelle zu Stelle* bewegten.
 (GE translation) ['from place to place']
 [...] under 1600-talet fick hästen en ny roll för att snabbt flytta artilleri *från ett ställe till ett annat* under strid.
 (SW translation) ['from one place to another']
- (2) They are phenomenally efficient at finding their way back to rewarding flower patches *over and over again*. (LEGS; EN original)
 Sie sind phänomenal effizient, wenn es darum geht, *immer wieder* zu lohnenden Wiesen zurückzukehren. (GE translation) ['ever again']
 De är utomordentligt skickliga på att *gång på gång* hitta tillbaka till givande blomsterängar. (SW translation) [Lit. 'time on time'; 'repeatedly']

When comparing the above originals with their translations, it is evident that there are different options available to translators. The English NPN binomial in (1) is translated into a parallel German NPN construction and a Swedish phraseological pattern (*från ett X till ett annat* ('from one X to another')). Interestingly, the Swedish translator could have chosen a parallel NPN binomial (*från ställe till ställe* ('from place to place')). The English 'ADV *and* ADV' construction in (2) is also rendered differently in the translations. The German translator here opts for the high-frequency phrase *immer wieder*, while the Swedish translator chooses an NPN construction.

In view of the variation observed above, the aim of this trilingual study is to explore distributions and uses of echoic time binomials in English, German and Swedish. The investigation focuses on the frequencies in originals and translations and the different lexical realizations or subtypes (i.e., NPN and 'X *and* X') in the three languages. Also, related recurrent patterns such as 'one X *at a time*' used as translation correspondences will be investigated, indicating the similarities and differences between the languages' preferences for different phraseological constructions.

2. Material and method

This study is mainly based on material from the Linnaeus University English-German-Swedish corpus (LEGS) (Ström Herold & Levin 2018, 2019) (version 0.3). Some additional material was collected from the English-Swedish Parallel Corpus

(ESPC) (Altenberg, Aijmer & Svensson 1999) to enable cross-genre comparisons. The trilingual and bidirectional structure of LEGS is shown in Figure 1.

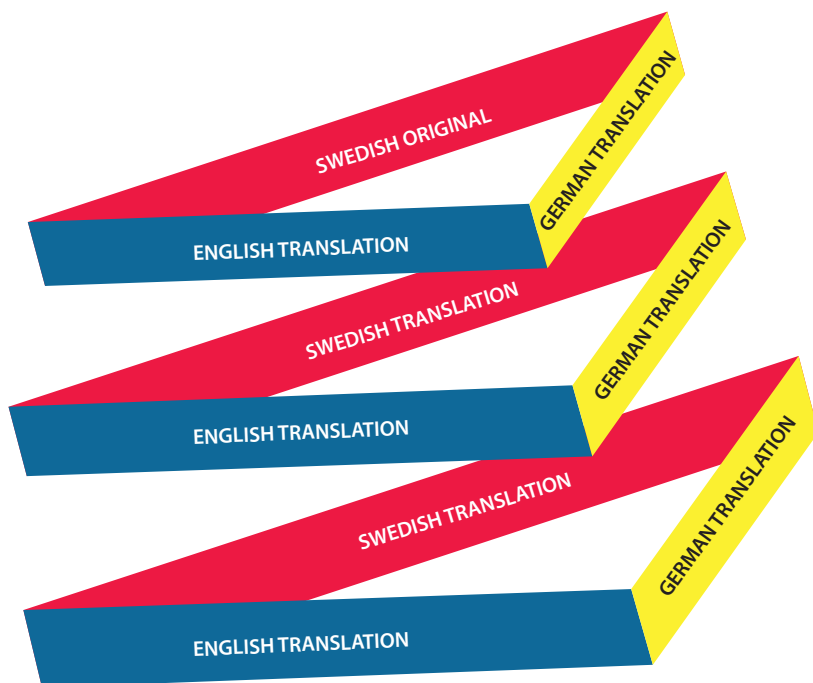


Figure 1. The structure of the Linnaeus University English-German-Swedish corpus (LEGS)

The corpus is balanced for all three languages, always including translations into two target languages, and consists of recently published (2000s) non-fiction texts from the subgenres popular history, popular science, biographies and self-help books. Each author and translator is represented just once. The present study covers ten English, eight German and nine Swedish books from LEGS. The original texts comprise 542,000 English, 370,000 German and 391,000 Swedish words.

The ESPC, in contrast, contains both fiction and non-fiction with texts from the 1970s to the 1990s amounting to 300,000 to 400,000 words per source-text corpus/genre (Altenberg, Aijmer & Svensson 1999). Since the ESPC search interface does not provide the option of searching for ‘XYX’ structures, we based our investigation on the original ESPC files, provided to us by Jarle Ebeling, University of Oslo.

First, a customized script retrieved all occurrences of ‘XYX’ in the originals and translations, i.e. instances where two identical items are separated by a third element. This procedure produced more noise (e.g., *I think I, as far as, stewing a whole chicken or chicken portions*) than relevant hits. In the next step, the irrelevant

hits were weeded out manually, and only instances with the structures NPN, 'ADJ and ADJ', 'ADV and ADV', 'NOUN and NOUN' and 'VERB and VERB' were kept.¹

Non-contiguous XYZ structures such as Jackendoff's (2008: 21) examples *gallon of paint after gallon of paint* and *day by miserable day* were not covered by the search, but some instances of this type were still retrieved when such structures occurred as translation correspondences for contiguous structures. Thus, the English phrase *(from) shell company to shell company* was included because it corresponds to the German original *(von) Briefkastenfirma zu Briefkastenfirma* ('(from) letterbox-company to letterbox-company'). As seen in these examples, binomials can occur in wider phraseological contexts than simple 'XYZ' structures: many NPN binomials are used in the '*from X to X*' construction (e.g., *from year to year*) and the English *over and over* and Swedish *om och om* are always followed by *again* and *igen* in the LEGS material.²

Since this study is restricted to echoic *time* binomials, we proceeded to omit all non-temporal instances. At this final weeding-out stage it became evident that a high proportion, about 75% (876 tokens) of all 'XYZ' binomials in LEGS, have temporal meaning. This proportion is in itself an important finding, supporting the idea of a strong connection of 'XYZ' constructions to dynamicity as also found by Knappe (2016). The non-temporal, stative instances that were not included in the analysis comprise both NPN constructions (*they lived side by side with the giant dinosaurs* (English original)) and 'ADV and ADV' (*durch und durch langweilig* ('through and through boring') (German translation)).

The echoic, repetitive, binomials that form the main interest of this paper have been largely overlooked in previous phraseological research. There are a few monolingual exceptions, mainly for English, that will be discussed in the following section. Systematic contrastive studies, however, are non-existent.

3. Echoic time binomials

Binomials have received considerable attention both from a synchronic and a diachronic perspective, mainly for English (Malkiel 1959; Gustafsson 1975; Kopaczkyk & Sauer 2017; Mollin 2017) but to some extent also for German (Lambrecht 1984).

-
1. Our material produced some rare instances of echoic trinomials (e.g., *We talk and talk and talk* (English translation from German)). These 'triplets' were counted as one instance each.
 2. Interestingly, Levin and Lindquist (2013: 14–15) note a decreasing use of *again*, so that in Present-Day American English the simple *over and over* is now more frequent than *over and over again*.

The focal points of the works mentioned have been (part-)synonymous, antonymous or complementary noun pairs linked by the coordinator *and* (e.g., *law and order*, *men and women* and *hand and foot*). Echoic, repetitive binomials, in contrast, are mostly discussed in passing (Malkiel 1959: 125–126, Gustafsson 1975: 9–10, 95), with a few exceptions such as Jackendoff (2008) on the English NPN, Lindquist and Levin (2009) on ‘N to N’, Knappe (2016) on the history of the English NPN, Ziem (2018) on the German NPN, and Levin and Lindquist (2013) on the ‘ADV and ADV’ construction. Despite this previous fragmentary coverage, the echoic type seems particularly varied and productive, showing the greater variety of links (Gustafsson 1975: 95).

The most in-depth study of echoic binomials, as we choose to call them, is Jackendoff’s comprehensive paper (2008) on the different subtypes of the English NPN construction. Jackendoff discusses a wide range of formal and functional features associated with the NPN construction. His study shows that the construction covers (semi)-productive patterns (frames) where the slots can be variably filled (see also Ziem (2018) for German data), albeit with certain formal restrictions on the repeated nouns such as: no mass nouns (**water after water*), no determiners (**the man for the man*) and no plurals (**men for men*). As for the links, Jackendoff argues that the pattern is productive with five prepositions: *after*, *by*, *for*, *to* and *upon* (*on*). Functionally, the NPN construction is, according to Jackendoff (2008: 9–10), usually used in adverbial function (*We examined the neighborhood, house by house*) but can also be found in premodifying position (*day-to-day progress*) (see also Levin & Ström Herold 2017). Jackendoff distinguishes five productive sub-constructions associated with standard senses:

- N *after* N (succession)
- N *by* N (succession or juxtaposition)
- N *for* N (comparison or exchange)
- N *to* N (juxtaposition or succession)
- N *upon* N (succession)

As can be seen, many of the senses expressed by NPN binomials relate to time, which is also the focus of our study. Interestingly, the NPN seems to have been originally dynamic in sense, only later developing static meanings such as juxtaposition (Knappe 2016). The temporal sense is also prevalent in the other construction type covered by this paper, namely the ‘X and X’ construction. Even less attention has been devoted to this type, the link *and* mostly being discussed in connection with non-echoic binomials (*men and women*, *back and forth*; Malkiel 1959; Mollin 2013). Levin and Lindquist (2013), however, who focus on the specific ‘ADV and ADV’ subtype, suggest that *on and on* and *over and over* (*again*) have been shifting from motion verbs (*ride on and on*; *turn over and over*) and now

express duration or repetition (*it went on and on; I told you over and over again*). Both the NPN and the ‘X and X’ binomial may serve as an intensifier, a common rhetorical function for reduplicative structures in general (Gustafsson 1975: 15), e.g. *walked and walked*. Jackendoff (2008: 14, 18) also notes the intensifying function of NPN binomials, arguing that, for instance, *day after day* expresses a higher degree of intensification than *every day*.

As a cross-linguistic observation, Gustafsson (1975: 10) suggests that binomials “may well be a universal feature in language”. Malkiel (1959) exemplifies binomials in a wide range of languages, but his study lacks comparisons of frequencies and distributions. Jackendoff mentions that the NPN construction is attested in other languages as well (e.g., Dutch, German, Japanese). While focusing on English, he suggests that “none of these other languages are likely to be exactly parallel” (2018: 8) but does not develop this further. Our study seeks to fill parts of this gap through a contrastive corpus-based account of time-related echoic binomials.

4. Results

Before moving on to discussions of the distributions of echoic binomials in the LEGS subcorpora, it is necessary to present the different types of binomials identified in the material. These are listed in Table 1. In the NPN type a preposition links two nouns, and in the remaining four, all being of the ‘X and X’ type, the coordinator *and/und/och* conjoins two identical items (adjectives (in the comparative), adverbs, verbs or nouns).

Table 1. Binomial types identified in the LEGS material

NPN	<i>day after day</i> / <i>Tag für</i> (‘for’) <i>Tag</i> / <i>dag efter</i> (‘after’) <i>dag</i> <i>pixel by pixel</i> / <i>Pixel für</i> (‘for’) <i>Pixel</i> / <i>pixel för</i> (‘for’) <i>pixel</i>
ADJ and ADJ	<i>younger and younger</i> / <i>yngre och yngre</i> (Sw.)
ADV and ADV	<i>over and over (again)</i> / <i>nach und nach</i> (‘after and after’) / <i>om och om (igen)</i> (‘over and over’)
VERB and VERB	<i>waited and waited</i> / <i>wartete und wartete</i> / <i>väntade och väntade</i>
N and N	<i>time and time (again)</i> <i>en och en</i> (Sw.) (lit. ‘one and one’; ‘one by one’)

A wide range of lexical items and lexical categories are attested in ‘XYX’ time binomials, but there are some differences between the three languages. The German

LEGS originals do not produce any binomials of the ‘ADJ *and* ADJ’ and the ‘N *and* N’ type. As will be seen below, ‘ADJ *and* ADJ’ binomials are rare in German in general, which suggests that German is less productive than English and Swedish, at least as regards the types of binomials.

As for lexical choices, it is clear that only a small portion of the English instances (24/109) contain items directly related to time such as *day*, *time* and *again*. The large majority of the instances therefore relate to time indirectly – their temporal, dynamic meaning being inferable from the immediate context. Thus, with verbs conjoined with *and*, the default interpretation is extension in time (e.g., *Florence waited and waited* (English original)), and if, using an NPN binomial, someone is described as rendering *an endless array of fonts [...] pixel by pixel on the screen* (English original) then this can only be interpreted as a process progressing over time. Notably, such examples align well with Knappe’s (2016) historical data on the English NPN construction which originally, as previously mentioned, expressed dynamic meanings.

This introductory section has shown the binomial types found in the three languages. Section 4.1 goes on to present the findings for the originals, starting with the overall frequencies, then moving on to the distributions of types and finally presenting the distributions of the links in the binomials.

4.1 Echoic time binomials in English, German and Swedish originals

The frequencies in the LEGS originals indicate that there are differences in the languages’ use of echoic time binomials. Figure 2 presents the raw numbers (in the boxes) and the normalized frequencies of NPN and ‘X *and* X’ binomials in English, German and Swedish originals.

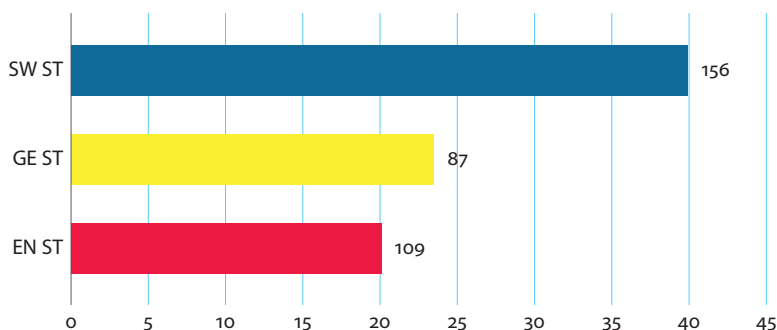


Figure 2. Frequencies of echoic time binomials in LEGS originals per 100,000 words

Compared to the English and German originals, echoic time binomials are significantly more frequent in the Swedish originals, while there is no notable difference between English and German.³ For Swedish originals, binomials are frequent across the non-fiction sub-genres and not an effect of individual texts skewing the data: four of the five highest frequencies for individual texts are found in Swedish originals. English originals show a different pattern with low frequencies and the least frequency variation between individual texts. A further interesting finding is that the observed differences between the texts in the corpus are not connected to any specific sub-genre. For example, both the highest and the lowest frequency for any texts were found in history books, the highest in the Swedish ones and the lowest in the German.

Although there are differences in the frequencies of binomials in the three subcorpora, there is no direct correlation between token frequencies and type-token ratios, at least not for NPN binomials. This is shown in Table 2, which gives the individual type-token ratios. The English original data produce both the lowest number of types and the lowest type-token ratio.

Table 2. Type-token ratios of NPN echoic time binomials in English, German and Swedish originals in LEGS

German	0.73	(37/51; 30 hapaxes)
Swedish	0.61	(63/103; 48 hapaxes)
English	0.57	(30/53; 18 hapaxes)

Thus, taken together, Figure 2 and Table 2 indicate that English neither uses particularly many echoic time binomials, nor is it more productive than German and Swedish in its use of NPN time binomials. In view of English being Jackendoff's (2008) inspiration for examining echoic binomials, this is perhaps unexpected.

The LEGS non-fiction data provide evidence of both high frequencies of some recurrent types and a high degree of creativity. The relatively low productivity of English NPN binomials is illustrated by four different types occurring in at least three of the ten English LEGS originals: *day after day*, *day to day*, *one by one* and *(from) time to time*. Swedish use in originals does not only differ from English and German as regards frequency, but also as regards types, as seen in Figure 3. German has a high degree of type productivity. Some of the more creative instances

3. The log-likelihood for each language comparison: SW-EN 30.9, $p = ***$; SW-GE 16.4, $p = ***$; EN-GE 1.18, $p = \text{ns}$. The log-likelihood test was preferred over the chi-squared test because the total number of observations (here, tokens) in the contingency table is very high and the log-likelihood test is arguably more robust than chi-squared in that scenario (Dunning 1993).

are exemplified in (3) and (4). In (3), the English translator chooses a similar NPN structure, while in (4) the translator resorts to another formulaic pattern (see further Section 4.3 below).

- (3) Beim Wieder-Aufrichten sollte man [...] den Rücken vom Becken aus nach oben *Wirbel für Wirbel* wieder aufrichten [...]
(LEGS; GE original) [‘vertebra for vertebra’]
To sit up again, [...] straighten up – *vertebra by vertebra* – beginning in the pelvis;
(EN translation)
- (4) „Solche Beziehungen wachsen wie hartes Holz [...] ganz langsam, *Jahresring für Jahresring*.“
(LEGS; GE original) [Lit. ‘year-ring for year-ring’]
“Such connections grow like hard wood [...] very slowly, *one ring every year*.”
(EN translation)

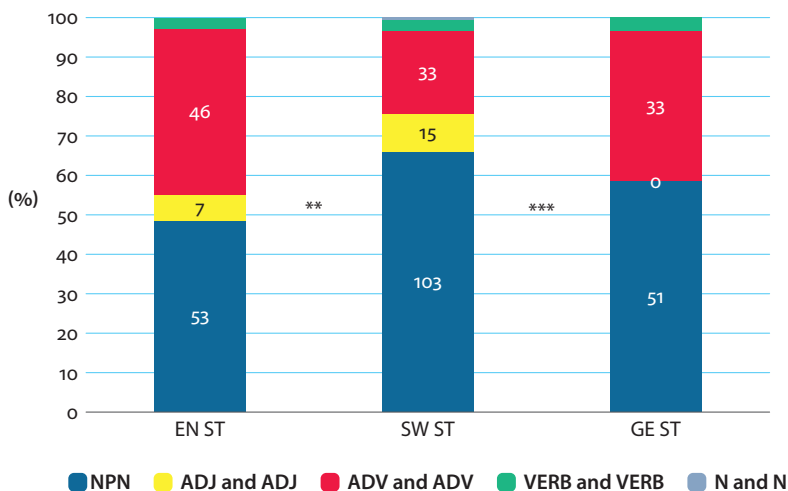


Figure 3. The distributions of echoic time binomial types in English, Swedish and German originals in LEGS

Swedish usage of echoic time binomials differs significantly from English and German, while there is no difference between the latter two.⁴ The Swedish

4. The results of a chi-squared test for independence for each significant language comparison: SW-EN, $\chi^2 = 14.3$, $df = 4$, $p = **$; SW-GE, $\chi^2 = 15.10$, $df = 4$, $p = ***$. Here and later, Bonferroni correction was applied to control the family-wise error rate, i.e., multiple comparison problem.

originals contain almost twice as many NPN binomials as English and German, and there are also more instances of 'ADJ *and* ADJ' in Swedish. As indicated in Table 2, the high frequency in Swedish originals is based both on a fairly frequent use of hapaxes, but also on some recurring types. The Swedish example in (5) shows three clustered hapaxes, illustrating how binomials can be used as a stylistic device. Similar examples of "stacked" binomials are noted by Lindquist and Levin (2009: 180), e.g., *Commuting [...], shoulder to shoulder, knee to knee with other riders.*

- (5) [...] företar nu ett dussin amerikanska jurister en utgrävning, sorterar och analyserar *dokument efter dokument, fakta efter fakta, brott efter brott.*
 (LEGS; SW original) ['document after document,
 fact after fact, crime after crime']
 [...] a dozen US lawyers are engaged in excavating, sorting, and analysing
document after document, fact after fact, crime after crime. (EN translation)

Among the recurrent Swedish binomials, the adverbial *gång på gång* stands out as the most frequent, occurring 19 times in seven (of nine) Swedish original texts. This binomial can be seen in (6) and refers to an event being repeated. Other binomial types recurring in the Swedish data include *år efter år*, occurring five times in four different texts (always translated into the identical English structure *year after year*) as exemplified in (7), and *kväll efter kväll* ('evening after evening'), *en efter en* ('one after one') and *steg för steg* ('step for step'), each being found in three texts. The high frequency of NPN binomials in Swedish is therefore a reflection both of a reliance on recurrent types and a large number of hapaxes.

- (6) Dessa personer dyker upp *gång på gång.*
 (LEGS; SW original) [Lit. 'time on time']
 These people pop up *time and again.* (EN translation)
- (7) Jag fattar inte hur han orkade *år efter år.*
 (LEGS; SW original) ['year after year']
 I don't know how he managed to keep it up *year after year.* (EN translation)

The 15 'ADJ *and* ADJ' instances occurring in Swedish originals comprise eight different types, one of which is exemplified in (8). Here the English translator opts for a 'simple' verb in spite of there being an available English construction parallel to the Swedish original (i.e., *grows thicker and thicker*).

- (8) Myelinlagret blir under utvecklingen *tjockare och tjockare [...]*
 (LEGS; SW original)
 [Lit. 'the-myelin-sheath becomes during the-development thicker and thicker']
 During development, this myelin sheath *thickens [...]* (EN translation)

The lack of ‘ADJ *and* ADJ’ instances in German LEGS originals reflects the low frequencies found in other corpora. Thus, for example, the 15-million-word DWDS-Kernkorpus (www.dwds.de) only produces a single instance of *größer und größer* (‘bigger and bigger’) but over a hundred of the favoured ‘immer COMP. ADJ’ alternative, *immer größer* (‘ever bigger’). As will be seen below, this construction is a highly frequent correspondence of English echoic binomials.

In all three source languages, the combined frequencies of ‘ADV *and* ADV’, ‘ADJ *and* ADJ’, ‘VERB *and* VERB’ and ‘NOUN *and* NOUN’ constructions make the coordinator *and/und/och* the most frequent link. This is shown in Table 3, which gives the raw numbers for *and/und/och* and the most frequent prepositions in each source language. Overall, the echoic binomial type seems to display a particularly varied set of links compared to non-echoic types, as also suggested by Gustafsson (1975: 95).

Table 3. The most frequent links for echoic time binomial types in English, German and Swedish originals in LEGS

EN source texts		GE source texts		SW source texts	
<i>and</i>	56	<i>und</i> (‘and’)	36	<i>och</i> (‘and’)	53
<i>to</i>	27	<i>für</i> (‘for’)	25	<i>efter</i> (‘after’)	35
<i>after</i>	12	<i>zu</i> (‘to’)	14	<i>på</i> (‘on’)	27
<i>by</i>	12	<i>um</i> (‘after/by’)	9	<i>till</i> (‘to’)	24
				<i>för</i> (‘for’)	15

The ability to combine elements from many different word classes is the reason for the high frequencies for the conjunctions *and/und/och*. As for prepositional links, there is more competition. Previous studies only provide limited frequency information for NPN constructions,⁵ but, impressionistically, Jackendoff (2008: 13) proposes that *by* is a more productive link than *to* and Gustafsson (1975: 95) suggests that *by* is the most frequent followed by *to* and *after*. These assumptions are not supported by our material.

A main difference between the three source languages in LEGS is that Swedish has slightly more recurring prepositions (four) than English and German (three each), reflecting the higher overall frequency of NPN in the Swedish texts. There

5. The three German case studies in Ziem (2018) are concerned with *an* (‘to/by’), *in* (‘in’) and *über* (‘over’) which do not reach top three among the echoic time binomials attested in LEGS.

are both similarities and differences in the types of prepositions in the source languages. In all three languages, *to* and its closest equivalents *zu* (Ge.) and *till* (Sw.) are virtually restricted to the ‘from X to X’ construction.⁶ Some prepositions are restricted to two languages: *after/efter* occur only in English and Swedish and ‘for’ only in German (*für*) and Swedish (*för*). The high frequency of NPN binomials in Swedish is thus both reflected in the recurrent use of individual prepositions and also in the use of slightly more preposition types.

This section has shown that echoic time binomials are most frequent in Swedish originals in LEGS and that this is mainly due to the NPN construction being particularly prevalent. English originals, in contrast, neither contain particularly many types nor tokens. Section 4.2 explores and compares the frequencies, forms and functions of echoic time binomials in English originals and translations.

4.2 Echoic time binomials in English originals and translations

The frequencies of echoic time binomials in English translations reflect the frequencies in the German and Swedish originals. This is seen when comparing raw numbers and numbers per 100,000 words between Figure 4 below and Figure 2 above. Figure 4 shows that English echoic time binomials are significantly more common in translations than in originals.⁷ The higher frequencies of binomials in the German and Swedish originals thus “shine through” in the English translations (cf. Hansen & Hansen-Schirra 2012). It can be noted that not all German and Swedish binomials are rendered as English binomials. However, there are even more binomials “added” in translations into English than “omitted”, which explains the higher number in the English translations. English translations from Swedish retain significantly more binomials while English translations from German add significantly more.⁸

Some clear tendencies emerge with the German and Swedish echoic time binomials not rendered as binomials in English. All German and most Swedish binomials not translated into English binomials are either of the NPN type (20 German and 21 Swedish tokens) or the ‘ADV and ADV’ type (24 German and 16

6. E.g., *from week to week*; *von Haus zu Haus* (‘from house to house’); *från gård till gård* (‘from farm to farm’). The only cases where *to* does not occur with *from* are the seven English premodifiers (e.g., *their day-to-day world*).

7. The log likelihood for each subcorpus comparison: EN ST–(GE >)EN 14.7, $p = **$; EN ST–(SW >)EN 36.7, $p = ***$; (GE >)EN–(SW >)EN 4.48, $p = *$.

8. Retained/omitted/added in German 43/44/89 and Swedish 112/44/67; overall $\chi^2 = 28.68$, $df = 2$, $p = ***$; retained $\chi^2 = 26.47$, $df = 1$, $p = ***$; omitted $\chi^2 = 1.3$, $df = 1$, $p = ns$; added $\chi^2 = 16.55$, $df = 2$, $p = ***$

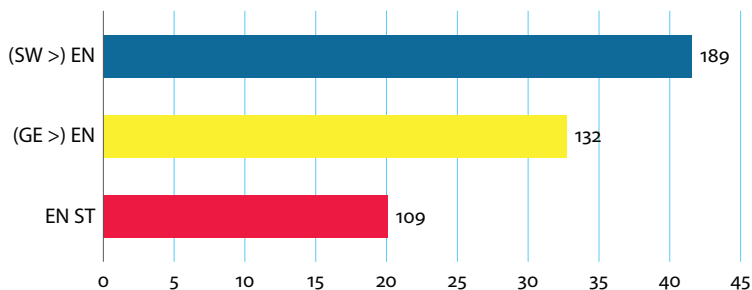


Figure 4. The frequencies of echoic time binomials in English source texts and translations in LEGS per 100,000 words

Swedish tokens). The most common correspondences are various “simple” English constructions (26 instances in translations from German and 20 in translations from Swedish binomials), such as the English - *ly* adverbial (e.g., German *wieder und wieder* ‘again and again’ > *repeatedly*; Swedish *mer och mer* ‘more and more’) and *steg för steg* ‘step by step’ > *gradually*). Another notable finding is that 11 Swedish binomials, mostly *då och då* ‘then and then’ and *gång på gång*, are translated into the English ‘X and Y’ pattern (*now and again*; *time and again* (as in (6)); *now and then*), indicating the strong interrelationship between echoic binomials and other recurrent patterns.

English binomials are added more often than they are omitted in translations from German and Swedish, which explains the higher frequencies of binomials in translations than in originals.⁹ Similar results of frequent units being even more frequent in translations than in originals are found in Ebeling and Ebeling (2017) who investigate 3-grams in originals and translations. The most frequent type of additions in English translations are of the ‘ADV and ADV’ type (42 instances from German originals and 45 instances from Swedish originals). This is mainly a result of the frequent German *immer* X and Swedish *allt* X ‘ever X’ typically being rendered as ‘ADV and ADV’ in English (37 German *immer* X; 30 Swedish *allt* X). Among these, the most prominent translation alternative is *more and more* (12 German and 22 Swedish instances). Two Swedish instances translated into *more and more* are given in (9), while (10) gives one of the many instances of the recurring German phrase *immer wieder* ‘ever again’ rendered as an English ‘ADV and ADV’ binomial.

9. More binomials are added in translations from German (89 instances) than from Swedish (67 instances).

- (9) [...] dataspelen skapar en alternativ värld där *allt fler* lägger *allt mer* tid.
(LEGS; SW original)

[Lit. 'ever more devote ever more time']

[...] computer games create alternative worlds into which *more and more* people are devoting *more and more* time. (EN translation)

- (10) *Immer wieder* unterstrich der Consultant, dass der Firmenverlauf genau seiner Vorhersage entsprach. (LEGS; GE original) ['ever again']
Again and again, the consultant stresses that the company's progress corresponds closely to his prediction. (EN translation)

As for NPN binomials, about equal numbers of binomials are added and omitted in English from both source languages (German: 20 NPN omitted and 29 added; Swedish: 21 NPN omitted and 18 added). The added instances mostly stem from various 'simple' instances in the German and Swedish originals, such as the simple Swedish prepositional phrase in (11) and the German adverb ending in *-weise* ('-wise') in (12).¹⁰ In these examples the translations seem to gain in intensity and dynamicity compared to the originals.

- (11) [...] polisinsatsen ökade *för varje år*.
(LEGS; SW original) [Lit. 'for each year']

[...] the police involvement increased *from year to year*. (EN translation)

- (12) *Schrittweise* tastet man sich durch die Dunkelheit, [...]
(LEGS; GE original) [Lit. 'stepwise']

You make your way through the darkness *step by step*, [...] (EN translation)

In Section 4.3, we discuss examples that show the exact opposite case, where the translation instead reduces the intensity and dynamicity of the original binomial (cf. Gustafsson (1975: 15) on the intensifying function of binomials). Loss in intensity is perhaps also the more expected pattern since translations are typically found to be lexically neutralized in relation to their originals (Baker 1996: 180).

Because of the dynamicity expressed by binomials, we expected high frequencies in narrative texts. To test this hypothesis, we compared the distributions of NPN binomials in LEGS to those found in the ESPC. This additional material

10. Interestingly, seven of the 29 added NPN binomials are translations of German *-weise* adverbs. The types include *schrittweise* (3 tokens) and *zentimeterweise* ['centimetre-wise'] (2 tokens). Although *-weise* is found to be translated into English binomials, the opposite is not the case. There were no instances of English binomials being translated into German *-weise*, which indicates that translators tend to avoid translating into more condensed structures.

comprises both original and translated fiction and non-fiction, which allowed further genre comparisons. This case study was restricted to binomials containing the three most frequent prepositions *after*, *by* and *to*. The frequencies of these (raw numbers and per 100,000 words) are given in Figure 5.

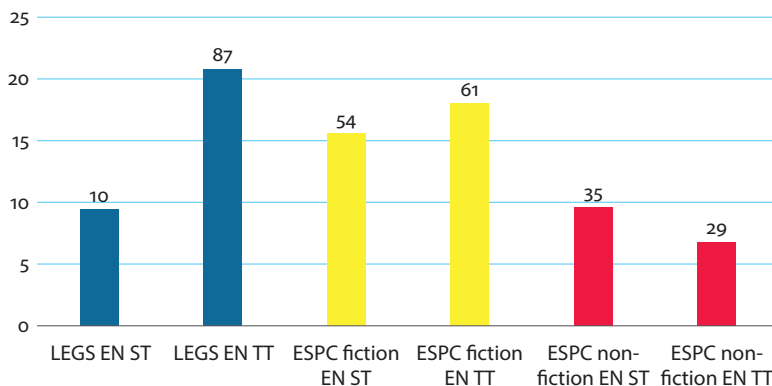


Figure 5. NPN echoic time binomials with *after/by/to* in English originals and translations from Swedish in LEGS and ESPC per 100,000 words

The figure shows that there are both similarities and differences between the corpora. The frequencies of NPN binomials are strikingly similar in LEGS and ESPC original non-fiction while, as expected, being significantly¹¹ more frequent in ESPC fiction than in the non-fiction corpora. This latter finding probably reflects the function and style of fiction texts compared to non-fiction. Fiction often involves narratives of events unfolding over time which would seem to favour echoic time binomials (e.g., *listening night after night to the never-ceasing moaning of the tide* (ESPC; English original fiction)). Moreover, some of the recurring strings are rather informal and therefore potentially more likely to occur in fiction than in non-fiction (e.g., *she had to find out everything herself, bit by bit* (ESPC; English translated fiction)).

In contrast to LEGS, the ESPC data do not produce any significant frequency differences between originals and translations. However, there are some similarities between the subcorpora. The ESPC fiction subcorpus shows a similar tendency to LEGS with more NPN binomials in translations than in originals, but this is not the case for the non-fiction texts. It is likely that the composition of the ESPC non-fiction subcorpus has influenced this result – the texts include, apart

11. ESPC English original fiction – LEGS English originals, LL 7.08, $p = **$; ESPC English original fiction – ESPC English original non-fiction, LL 5.48, $p = *$

from popular science texts, also specialized genres lacking equivalents in LEGS, such as EU documents and political speeches.

Lexical impoverishment has been suggested as a translation universal (Baker 1996). Because of this, we also investigated the type-token ratios of NPN binomials in LEGS and the two ESPC subcorpora. The translations from Swedish produce a higher type-token ratio (0.66) than English originals (0.55) with binomials containing *after*, *by* and *to* in LEGS, but this is likely due to there being almost twice as many tokens in the translations. The expected trend of higher ratios in originals than translations was, however, found for both fiction and non-fiction in the ESPC where the token ratios are more equal across originals and translations. Thus, at least for the ESPC data, there is some evidence of translationese for NPN binomials (cf. Gellerstam 1986).

Returning to the findings in LEGS, there are, as seen above in Figure 3, significant differences in the distributions of the different types of echoic time binomials in English, German and Swedish originals. In translations into English, there are many cases of binomials being added or omitted, and because of these changes in translations, the proportional distributions of binomial types in English originals do not differ from those found in translations from German and Swedish, as seen in Figure 6.

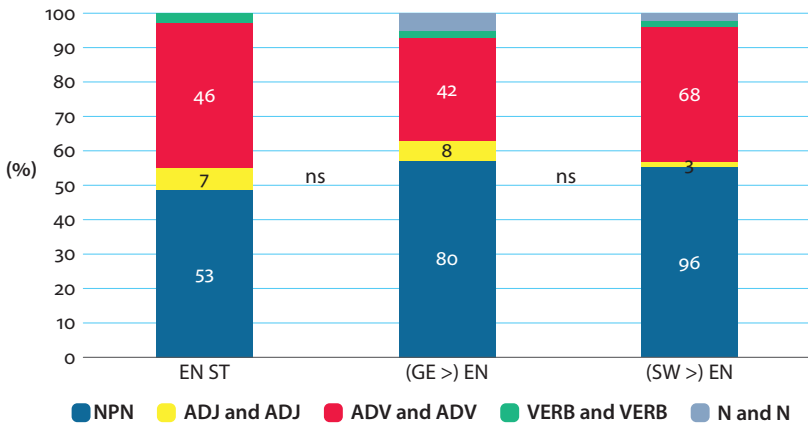


Figure 6. Distributions of types of echoic time binomials in English originals and translations in LEGS

Figure 6 thus shows that the translators manage to balance out the different proportions of the five types (NPN, 'ADV and ADV', 'ADJ and ADJ', 'VERB and VERB', 'NOUN and NOUN') in Swedish originals and English originals. As seen in Figure 4, however, the overall higher frequencies in German and Swedish originals nevertheless shine through in the translations into English.

As also seen in Figure 6, binomials linked with the coordinator *and* are quite frequent in both English originals and translations. The raw frequencies of *and* and the prepositions *after*, *by* and *to* are given in Table 4. As with the German and Swedish originals in Table 3 above, the most common link in English translations is *and*.

Table 4. The most frequent links in English originals and translations

Link	EN ST	(GE >) EN	(SW >) EN	Example
<i>and</i>	56	72	83	<i>on and on</i>
<i>to</i>	27	21	25	<i>door to door</i>
<i>after</i>	12	10	36	<i>time after time</i>
<i>by</i>	12	25	15	<i>mile by mile</i>

As indicated above, *and* is more frequent in translations than in originals. This is largely due to translators adding many ‘ADV *and* ADV’ constructions that are used as correspondences for *immer/allt* X. In contrast to the link *and*, there are notable differences in the frequencies of the prepositions, *to* being the most common in English originals, *by* in translations from German and *after* in translations from Swedish. These differences are partly reflections of the prepositions used in the German and Swedish originals. The most notable example of this is *after*, which is the third most frequent preposition in translations from German and the most frequent in translations from Swedish. As discussed above, English *after* lacks a productive German equivalent link (*nach*), while the Swedish equivalent (*efter*) is the most frequent in Swedish originals. These preferences, however, are not enough to create significant differences in the use of prepositions between English originals and translations.¹² However, the differences between translations from German and translations from Swedish are highly significant.

The comparisons in this section have shown that translations into English contain more echoic time binomials than English originals, a result largely reflecting the higher frequencies of such constructions in German and Swedish originals. The differences in the types (NPN, ‘ADV *and* ADV’) of English binomials in originals and translations are much less noticeable. The final section will have a reversed focus, targeting the constructions used in German and Swedish

12. Neither are there significant difference in preposition usage between originals and translations in ESPC: ESPC fiction originals (*after*: 12, *by*: 19, *to*: 23) and translations (*after*: 19, *by*: 23, *to*: 23); ESPC non-fiction originals (*after*: 5, *by*: 18, *to*: 12) and translations (*after*: 7, *by*: 13, *to*: 9). The differences between German and Swedish: $\chi^2 = 14.85$, $df = 2$, $p = ***$

translations. One important finding in this section is that there is a stronger correlation between English and Swedish binomials than between English and German.

4.3 German and Swedish correspondences to English echoic time binomials

Four correspondence categories of English echoic time binomials were identified in the German and Swedish translations: i) binomials, ii) phraseological patterns, iii) simple constructions and iv) zero constructions. These are exemplified in (13) to (20).

The binomials category includes both instances where the same type of structure is used in the original as in the translation – as in (13), where an English NPN is rendered as an NPN binomial in Swedish – and also cases of ‘type shifts’ – as in (14) and (15) – where one category of original binomial is translated into another. In (14) the English NPN *one by one* is translated into a German ‘ADV and ADV’, and in (15) the English ‘ADV and ADV’ into a Swedish NPN.

Binomials

- (13) Perhaps traces of the pheromones linger *from year to year*, [...] (LEGS; EN original)
Kanske dröjer sig feromonspår kvar *från år till år*, [...] (SW translation) [Lit. ‘from year to year’]
- (14) By the 1980s they were known only in a handful of places, and *one by one*, those populations disappeared. (LEGS; EN original)
In den 1980ern kannte man sie nur noch an einigen wenigen Orten, und auch diese Populationen sind *nach und nach* verschwunden. (GE translation) [Lit. ‘after and after’; ‘gradually’]
- (15) And it is this neglect that, *over and over again*, turns natural disasters into unnatural catastrophes. (LEGS; EN original)
Och det är denna vanvård som *gång på gång* gör naturkatastrofer till onaturliga katastrofer. (SW translation) [Lit. ‘time on time’; ‘repeatedly’]

There were twelve occurrences of such type shifts in translations, three in German and nine in Swedish. Of these, two involved English NPN being translated into German *nach und nach* and five instances of *over and over (again)* rendered as the Swedish NPN *gång på gång* (see also (2) above). These kinds of cross-linguistic correspondences show that there are strong parallels between the different types of echoic time binomials investigated in the present paper.

The patterns category includes various phraseological patterns with empty slots (excluding the ‘XYX’ types). The most frequent type, *immer/allt* X, exemplified in (16), was discussed above. One of the many rarer patterns, Swedish *en/ett* X *i taget*, is shown in (17).

Patterns

- (16) *Many bees evolved longer and longer* tongues to make it easier for them to reach nectar hidden within flowers; (LEGS; EN original)
 Viele Bienen entwickelten *immer längere* Rüssel, um besser an den in der Blüte verborgenen Nektar zu gelangen; (GE translation) ['ever longer']
 Många biarter utvecklade *allt längre* tungor för att lättare kunna nå blommornas gömda nektar [...]; (SW translation) ['ever longer']
- (17) Documents should not lurch *line by line* as you scroll through them, but instead should flow. (LEGS; EN original)
 dokumenten skulle inte hacka fram *en rad i taget* när man scrollade, utan flyta fram. (SW translation) ['one line at a time']

The simple category is illustrated in (18) and (19). It covers instances where the correspondences are lexically “simple” constructions, i.e. not phraseological patterns. In (18) both translators opt for simple adjectives for the English ‘ADV *and* ADV’ binomial. For either target language, it is difficult to suggest equivalent binomials that could be used as correspondences. The same holds true for (19) where the binomial in the English original functions as a complex premodifier (Levin & Ström Herold 2017), an option which is precluded in the target languages (**dag-för-dag-omvårdnad*).

Simple

- (18) The list of benefits goes *on and on*. (LEGS; EN original)
 Die Liste der positiven Effekte ist *schier endlos*.
 (GE translation) ['the list [...] is almost endless']
 Listan över fördelar kan göras *lång*.
 (SW translation) ['the list [...] can be made long']
- (19) Their *day-to-day* welfare was attended to by other people, [...] (LEGS; EN original)
 Um das *tägliche* Wohlergehen der Kinder kümmerten sich andere, [...] (GE translation) ['daily']
 Deras *dagliga* omvårdnad sköttes av andra personer, [...] (SW translation) ['daily']

As previously mentioned, Gustafsson (1975: 15) and Jackendoff (2008: 14, 18) argue that binomials are intensifiers. Thus, the ‘simple’ translations above can be said to be examples of ‘de-intensification’ or neutralization in translation, which, as seen above, has been proposed as a translation universal (Baker 1996: 180).

Zero instances are the rarest category. In these, the translators opt for excluding the meaning of the binomial, i.e. an implicitation strategy (Ingo 2007: 124).

Translators typically omit the binomial when the information can be readily inferred from the context. An example is given in (20), where *one by one* is near-synonymous to the phrase *individual stores*. The translator renders this as *persönlich* ('in person'), which we argue, is more closely equivalent to *individual* than the omitted binomial.

Zero

- (20) [...] Jobs [...] seemed content to peddle his product to individual stores *one by one*. (LEGS; EN original)

[...] dass Jobs [...] zufrieden zu sein schien, seine Produkte den Elektronikgeschäften persönlich anzubieten Ø.

(GE translation) ['in person offer his products to electronics-stores']

The German and Swedish correspondences in translations from English are presented in Figure 7.

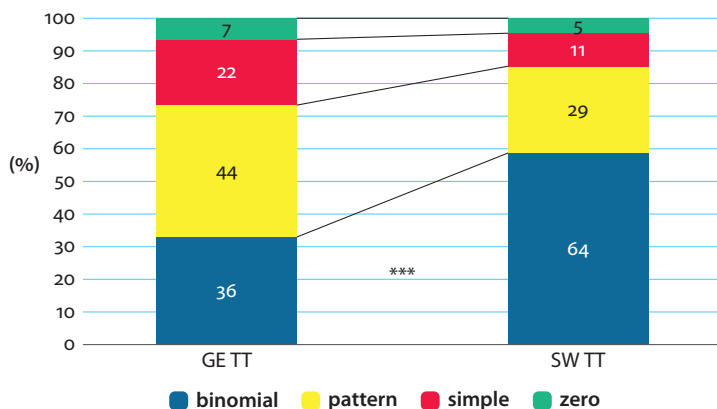


Figure 7. German and Swedish translation correspondences of English original echoic time binomials in LEGS

There is a highly significant difference for binomials as translation correspondences – Swedish translators use binomials in more than half the instances while this applies to only a third of the German. Thus, there is a higher level of correspondence between English and Swedish binomials than between English and German ones. German instead uses more of the other three categories – patterns, simple and zero. An example of a German zero instance is given in (20) above, and below, in (21), is an instance where the German translator makes use of a simple construction for English *over and over again* while the Swedish resorts to an ADV and ADV construction parallel to the English one. For German, it would have been possible to use the binomial *wieder und wieder* ('again and again'), but instead the iterative meaning referring to the researchers' activities of the original is reinterpreted as a continuative event referring to the children's behaviour.

- (21) [...] the team observed what each child did, *over and over again*, [...]
 (LEGS; EN original)
 [...] beobachtete das Team, wie sich jedes Kind bei sozialen Interaktionen
 [...] *dauerhaft* verhielt.
 (GE translation) ['how each child continuously behaved']
 [...] observerade teamet vad varje barn gjorde, *om och om igen*, [...]
 (SW translation) ['over and over again']

The German predilection for patterns is exemplified in (22), which is one of three instances where English *one by one* is translated into the German pattern *einer nach dem anderen* ('one after the other'). In this case, there is no equivalent binomial available in German (**einer nach einem*). It is also interesting to note that there is a functional change in that the English binomial adverbial is transposed into a German pattern as a subject.

- (22) Everybody got the idea they were working very hard for Robert's farm, and
one by one they started to leave. (LEGS; EN original)
 Jeder hatte die Vorstellung, dass er für Roberts Farm sehr hart arbeitete,
 und *einer nach dem anderen* verließ die Farm.
 (GE translation) ['one after the other']
 Det började gå upp för folk att de slet väldigt hårt för husbonden Robert,
 och de gav sig av *en efter en*. (SW translation) ['one after one']

As indicated in Section 4.2, *immer/allt X* is by far the most frequent phraseological pattern used in German and Swedish as correspondences of English binomials. *Immer X* is slightly more common in German than *allt X* in Swedish, but if we disregard the *immer/allt X* instances there are both more construction types (6 vs. 3) and tokens (11 vs. 6) in the pattern category in Swedish compared to German translations. Among the recurrent patterns found in German and Swedish are 'X and Y' (*sway from side to side* > *pendeln* [...] *hin und her* (Ge.); *svävar av och an* (Sw.) ('back and forth')) and 'from X to Y' as in the Swedish translation in (1) above.

Above we discussed the German and Swedish translation correspondences of English echoic time binomials. In the following we turn to the German and Swedish original structures translated into English binomials. These are shown in Figure 8.

As with the translations in Figure 7, Figure 8 indicates a significantly stronger correlation between English and Swedish binomials than between English and German. This difference is largely an effect of Swedish NPN binomials (82 occurrences) more often being rendered as English binomials than German ones (31 occurrences). Most of these Swedish instances (75) were rendered as NPN

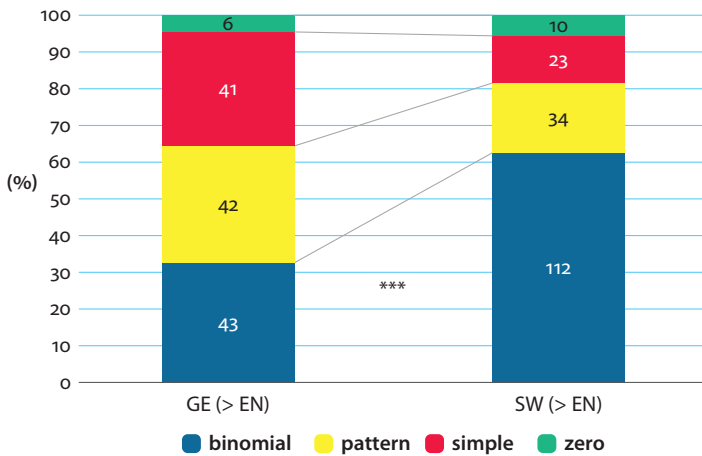


Figure 8. Original German and Swedish correspondences of translated English echoic time binomials in LEGS

binomials in English.¹³ Two of these are shown in (23) where the productive ‘*from X to X*’ construction is readily transferred from one language to another. For German originals, only 30 NPN binomials were rendered as English NPNs. One of these instances is given in (24), which contains the recurrent types *Stück für Stück* and *bit by bit*.

- (23) Vakenhetsgraden kan variera *från sekund till sekund* och *från timme till timme*. (LEGS; SW original) [‘from second to second; from hour to hour’]
Levels of arousal can vary *from second to second* and *from hour to hour*.
(EN translation)
- (24) Es besteht aus besonders komplizierten Zuckerketten, die *Stück für Stück* auseinandergebaut werden müssen. (LEGS; GE original) [‘bit by bit’]
Such bread contains particularly complex sugar chains, which have to be broken down *bit by bit*.
(EN translation)

Also, in translations into English, *immer/allt X* predominates in the pattern category (37 of 42 instances in German; 30 of 34 instances in Swedish). Among the simple types, we have already, in Section 4.2, commented on the seven instances of German *-weise* translated into English binomials. The simple constructions translated into English binomials include many instances of simple adverbials, such as those in (25) and (26).

13. Apart from *gång på gång* being translated as ‘ADV and ADV’ 5 times (*again and again* 4; *over and over* 1) and ‘N and N’ twice (*time and time again*).

- (25) Gör om försöket *många gånger*, [...] (LEGS; SW original) ['many times']
Repeat the experiment *again and again*, [...] (EN translation)
- (26) Man hör uttrycket *ibland*, entomologer emellan:
(LEGS; SW original) ['sometimes']
I hear the expression *from time to time* among entomologists.
(EN translation)

These examples also illustrate a type of intensification in translation. Although, as discussed above, de-intensification, or neutralization, may be assumed to be a more common (universal) translation strategy, Figures 7 and 8 indicate that intensification and de-intensification are about equally common in translations from and into English with echoic binomials in LEGS.

This section has provided two main findings on German and Swedish binomials in relation to English: (i) a large majority of English echoic time binomials are translated either into or from German and Swedish binomials or phraseological patterns, and (ii) there is a stronger correlation between Swedish and English binomials than between German and English ones. The latter finding is partly related to the previously noted German predilection for derived adverbials with *-weise* and its greater reliance on patterns, most notably 'immer COMP.ADJ'.

5. Conclusion

Although previous studies have touched upon the use of echoic binomials from a cross-linguistic perspective, none has attempted a systematic comparison of actual corpus data. This study therefore provides a much needed broadening of our knowledge about such reduplicative structures.

The guiding assumption in this paper has been that multi-directional translation data can produce insights that are difficult to obtain with other kinds of material. A first finding from our balanced trilingual corpus LEGS is that echoic time binomials are pervasive in all three languages, NPN binomials being the most frequent type in all three source languages, followed by the 'ADV and ADV' type. While the 'ADV and ADV' structure produces few types (e.g., *again and again*, *nach und nach*, *om och om*) the NPN is more productive, including recurrent types such as *day after day*, but also creative hapaxes such as the German (*von*) *Briefkastenfirma zu Briefkastenfirma* and *Wirbel für Wirbel*. However, our findings also show that there are clear differences in distribution and variation. The Swedish originals exhibit the highest frequency of echoic time binomials, largely due to the productivity of the NPN construction (as also seen in the wide set of links), compared to the English and German originals. In fact, the English originals exhibit

the least variation in the use of NPN binomials, a possibly surprising result considering that Jackendoff (2008) highlights the versatility of the English NPN construction. Our data clearly suggest that this versatility extends to other languages as well, but further studies are called for to explore the usage patterns of echoic binomials in (non-)Germanic languages.

When comparing English originals and English translations, the frequencies of binomials in translations into English reflect the higher frequencies in the German and Swedish originals, which suggests that the preferences in the originals shine through in the translations, a commonly attested translation phenomenon. Our data also reveal substantial differences between the German and Swedish translations of English echoic time binomials, indicating that there is a stronger correlation between Swedish and English binomials than between German and English. The main finding here is that Swedish translations resort to binomials in more than half the instances while German translations only do so a third of the time and more often use patterns or simple constructions as translation correspondences. The fact that echoic binomials often correlate with recurrent phraseological patterns in other languages is perhaps not surprising in view of Stubbs' (2007: 97) suggestion that recurrent meanings tend to be expressed by recurrent patterns.

References

- Altenberg, Bengt, Aijmer, Karin & Svensson, Mikael. 1999. *The English-Swedish Parallel Corpus (ESPC): Manual*. Lund: Department of English, Lund University.
- Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager* [Benjamins Translation Library 18], Harold Somers (ed.), 175–186. Amsterdam: John Benjamins. <https://doi.org/10.1075/btl.18.17bak>
- Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19(1): 61–74.
- Ebeling, Signe Oksefjell & Ebeling, Jarle. 2017. A functional comparison of recurrent word-combinations in English original vs. *translated texts*. *ICAME Journal* 41: 31–52. <https://doi.org/10.1515/icame-2017-0002>
- Gellerstam, Martin. 1986. Translationese in Swedish novels translated from English. In *Translation Studies in Scandinavia*, Lars Wollin & Hans Lindquist (eds), 88–95. Lund: CWK: Gleerup.
- Gustafsson, Marita. 1975. *Binomial Expressions in Present-day English*. Turku: University of Turku.
- Hansen, Sandra & Hansen-Schirra, Silvia. 2012. Grammatical shifts in English–German noun phrases. In *Cross-linguistic Corpora for the Study of Translations: Insights from the Language Pair English–German*, Silvia Hansen, Erich Steiner & Stella Neumann (eds), 133–145. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110260328.133>

- Ingo, Rune. 2007. *Konsten att översätta: Översättandets praktik och didaktik*. Lund: Studentlitteratur.
- Jackendoff, Ray. 2008. Construction after construction and its theoretical challenges. *Language* 84(1): 8–28. <https://doi.org/10.1353/lan.2008.0058>
- Knappe, Gabriele. 2016. *Wunder æfter wundre*: The earliest history of the NPN construction in English. Paper presented at ISLE 4, in Poznan.
- Kopaczky, Joanna & Sauer, Hans. 2017. Defining and exploring binomials. In *Binomials in the History of English: Fixed and Flexible*, Joanna Kopaczky & Hans Sauer (eds), 1–23. Cambridge: CUP. <https://doi.org/10.1017/9781316339770.001>
- Lambrecht, Knud. 1984. Formulaicity, frame semantics, and pragmatics in German binomial expressions. *Language* 60(4): 753–796. <https://doi.org/10.1353/lan.1984.0004>
- Levin, Magnus & Lindquist, Hans. 2013. *Like I said again and again and over and over*: On the ADV1 and ADV1 construction with adverbs of direction in English. *International Journal of Corpus Linguistics*, 18(1): 7–33. Reprinted (2015) in *Current Issues in Phraseology*, Sebastian Hoffmann, Bettina Fischer-Starcke & Andrea Sand (eds), 7–33. Amsterdam: John Benjamins. <https://doi.org/10.1075/ijcl.18.1.04lev>
- Levin, Magnus & Ström Herold, Jenny. 2017. Premodification in translation: English hyphenated premodifiers in fiction and their translations into German and Swedish. In *Cross-linguistic Correspondences. From Lexis to Genre*, Thomas Egan & Hildegunn Dirdal (eds), 149–175. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.191.06lev>
- Lindquist, Hans & Levin, Magnus. 2009. The grammatical properties of recurrent phrases with body part nouns: The N_1 to N_1 pattern. In *Exploring the Lexis-Grammar Interface* [Studies in Corpus Linguistics 35], Ute Römer & Rainer Schulze (eds), 171–188. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.35.12lin>
- Malkiel, Yakov. 1959. Studies in irreversible binomials. *Lingua* 8: 113–160. [https://doi.org/10.1016/0024-3841\(59\)90018-X](https://doi.org/10.1016/0024-3841(59)90018-X)
- Mollin, Sandra. 2013. Pathways of change in the diachronic development of binomial reversibility in Late Modern American English. *Journal of English Linguistics* 41(2): 168–203. <https://doi.org/10.1177/0075424213478980>
- Mollin, Sandra. 2017. Developments in the frequency of English binomials, 1600–2000. In *Binomials in the History of English: Fixed and Flexible*, Joanna Kopaczky & Hans Sauer (eds), 281–295. Cambridge: CUP. <https://doi.org/10.1017/9781316339770.015>
- Ström Herold, Jenny & Levin, Magnus. 2018. English supplementary *ing*-clauses and their German and Swedish correspondences. In *Corpora et Comparatio Linguarum: Textual and Contextual Perspectives*, Signe Oksefjell Ebeling & Hilde Hasselgård (eds). *Bergen Language and Linguistics Studies* 9(1): 115–138. <https://doi.org/10.15845/bells.v9i1.1522>
- Ström Herold, Jenny & Levin, Magnus. 2019. *The Obama presidency, the Macintosh keyboard and the Norway fiasco*: English proper noun modifiers and their German and Swedish correspondences. *English Language and Linguistics* 23(4): 827–854. <https://doi.org/10.1017/S1360674319000285>
- Stubbs, Michael. 2007. An example of frequent English phraseology: Distributions, structures and functions. In *Corpus Linguistics 25 Years on*, Roberta Facchinetti (ed.), 89–105. Amsterdam: Rodopi.
- Ziem, Alexander. 2018. *Tag für Tag Arbeit über Arbeit*: Konstruktionsgrammatische Zugänge zu Reduplikationsstrukturen im Deutschen. In *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*, Kathrin Steyer (ed.), 25–48. Tübingen: Narr Francke Attempto.

CHAPTER 7

Just a moment

Brief times in English and Norwegian

Thomas Egan & Siri Frst Skogmo

Inland Norway University of Applied Sciences

This chapter investigates two temporal lexemes, English *moment* and Norwegian *yeblikk*, which may encode either a point in time ('x happened at time y') or an interval of time ('x lasted for time y'). The data for the study comprise all instances of the two lexemes in the English and Norwegian original language texts and translations in the English–Norwegian Parallel Corpus. While both lexemes are commonly used to encode time intervals, *moment* is almost twice as likely as *yeblikk* to refer to a point in time, suggesting that *moment* is somewhat more polysemous. This impression is reinforced by the evidence of the translations, where *moment* is more likely to be used to translate *yeblikk* than vice versa.

1. Introduction

This chapter presents the results of an investigation into the coding of points in time and brief intervals of time by English *moment* and its closest Norwegian correspondent *yeblikk* ('blink/wink of an eye'). The data for the study is taken from both the fictional and non-fictional texts in the bidirectional English–Norwegian Parallel Corpus (ENPC, see Johansson 2007a).

Examples (1) and (2) show both *moment* and *yeblikk* coding a point in time and a short temporal interval, respectively.

- (1) *At that moment* the bell in the corridor sounded for the end of class. (RD1)¹
I samme yeblikk ringte skoleklokka ute i korridoren og fortalte at
timen var over. (RD1T)

1. (RD1) indicates the provenance of the token, 'RD' being the initials of the author. (RD1T) serves to indicate a translation of the same text, 'T' standing for 'translation'. All tokens are cited in both original and translated forms, with the addition of a short gloss by the authors in cases where the translation is divergent in form or function from the source.

(2) They were quiet *a moment*.

(AT1)

De var stille *et øyeblikk*.

(AT1T)

In (1) the phrases *at that moment/i samme øyeblikk* pin down the exact point in time at which the event of bell ringing occurred. In (2) the phrases *a moment/et øyeblikk* convey the information that the state of quietness was of brief duration.

The present study aims to answer three research questions.

1. What are the similarities and differences between the distribution of *moment* and *øyeblikk* in English and Norwegian respectively?
2. What multi-word combinations do they both occur in?
3. What are the main translation correspondences of the two lexemes?

Moment is one of only a small handful of English lexemes, and *øyeblikk* one of a handful of Norwegian lexemes, to be commonly used to encode both time points and intervals. The distinction between points in time and intervals of time is pervasive in temporal expressions and is germane to considerations of lexical aspect/aktionsart (Vendler 1957). It is explored in more detail in Section 2. Section 3 contains a description of the corpus used for the present study. Section 4 presents an overview and contrastive analysis of tokens of *moment* and *øyeblikk* in the source language texts in the ENPC. In Section 5 we consider the translation correspondences of the two lexemes, while Section 6 is devoted to their non-correspondences, exemplified by multi-word combinations in both the original texts and their translations. Finally, Section 7 contains a summary and conclusion.

2. Points and intervals

According to Haspelmath (1997: 1), time is one of “the two most important basic conceptual domains of human thinking”, the other being space. All languages contain some means of locating situations with respect to time. Without such a means, it would be difficult to conceive of human coexistence. As Malinowski put it, over 90 years ago:

A system of reckoning time is a practical, as well as a sentimental, necessity in every culture, however simple. Members of every human group have the need of coordinating various activities, of fixing dates for the future, of placing reminiscences in the past, of gauging the length of bygone periods and of those to come.

(Malinowski 1927: 203)

On the subject of reckoning time, Munn (1992: 102) writes: “Strictly construed, ‘time-reckoning’ refers to the use of selected cultural categories, or contingent events [...] to ‘tell time’ – to ask ‘when’ something happened, will or should

happen – and to ‘measure’ duration – to ask ‘how long’ something takes, or to ‘time’ it.” These two forms of reckoning, used to communicate either when a situation occurred or how long one lasted, are investigated by Haspelmath (1997) in a study of NP-based temporal adverbials in 53 languages. Haspelmath labels the two ‘simultaneous temporal location’ (i.e. *x* happened at a certain time) and ‘temporal extent’ (i.e. *x* lasted for a certain time). The noun *time* is itself used in English NP-based expressions for both senses, *for a time/at a time*, labelled ‘time as interval’ and ‘time as moment’ respectively by Evans (2004). *Moment* is also commonly used in English in both senses, although in its case the time interval sense normally, though not invariably, denotes a very brief period.

Of the two notions, time as a point, and time as an interval or limited period, the former is conceptually simpler since the latter presumes a plurality of points. According to Evans (2004: 16), a “large literature gives rise to the view that humans directly perceive and experience duration and simultaneity”. They do so by means of a perceptual code which “consists of a temporal interval characterised by the correlated oscillation of neurons, which lasts for a short period of time. This phenomenon has been termed a PERCEPTUAL MOMENT” (Evans 2004: 22). Despite the fact that Evans uses the term ‘interval’ in his definition, he is referring here to ‘time when’ rather than ‘time during which’. The latter concept requires the presence of at least two perceptual moments, as well as the perception that these are related in sequence. As Evans himself put it, “a perceptual moment alone is unable to furnish the experiences of temporality. For this to emerge a rudimentary memory is required which relates the perceptual moment just experienced with the updated perceptual moment” (Evans 2004: 26). The subjective perception of the relationship between two points is central to the notion of time interval, since it is perfectly possible to conceive of a plurality of temporally unrelated perceptual moments, as in (3).

- (3) In their panicky *moments* they were sure they would bump into things.

(RDA1)

I øyeblikk av panikk var de sikre på at de ville komme til å støte borti ting.

(RDA1T)

The conceived extent of a perceptual moment depends not only on its actual length but also on the perspective in which it is viewed. According to Evans, “From the perspective of the Moment Sense, past intervals held in memory may, due to the erosion of episodic memory, lose their durational significance, and accordingly become ‘point like’” (Evans 2004: 129). Although the loss of durational significance may be more likely when it comes to past situations, it is perfectly possible to conceive of a present situation as instantiating an extended point rather than an interval. Example (4) is a case in point.

- (4) Authorship is somewhat out of fashion *at the moment*. (HB1)
 Forfatterskap er ikke på moten *for tiden*.
 (HB1T). ['...for the (present) time.']

The expression *at the moment* in (4) does not carry the implication that the state described is of limited duration. Nor does it exclude the possibility of the state referred to coming to an end. However, one could perfectly well append 'and will forever be so' to the predication as it stands. This would not be possible in a *for* construction: *for the moment*, as in (5), implies that the situation predicated is destined to come to an end.

- (5) Let us leave her *for the moment* and go back to Matilda and her first day in Miss Honey's class. (RD1)
 Nå skal vi glemme henne *et øyeblikk* og vende tilbake til den første dagen Matilda var på skolen og kom i klassen til frøken Honey.
 (RD1T) ['... a moment ...']

The phrase *for the moment* in (5) actually tells us both when something (*leaving her*) happened and informs us that this situation will only last for a certain time. It may thus be said to encode both a point and an interval. The point reading is triggered by the definite article, the interval sense by the preposition. The point reading is however redundant. If we were to omit the phrase, we would still be able to deduce 'time when' from the imperative verb. One would, on the other hand, lose the implication that the situation is of temporary duration. We therefore decided to classify tokens of *for the moment* as encoding an interval in our classification. Note, by the way, that the interval sense is also prioritised in (5) by the translator, who uses the indefinite article.

Given that *at* temporal constructions in English encode points in time, and *for* temporal constructions intervals, the questions arise as to how often and with what nominals they do so. In general clock and calendar nominals such as *second*, *minute*, *hour*, *day*, *week* and *year* occur regularly as complements of *for* (sense 28a in the OED), but very rarely with *at* (sense 29a in the OED). Two exceptions are *minute* and *hour*. Examples where these occur with the definite article are given in (6) and (7), both of which are from the original (2001) version of the British National Corpus (BNC).

- (6) We play dominoes more than cards *at the minute* though. (BNC KCP)
 (7) Margaret sent after Richard only to learn that he had died *at the hour* she was afflicted. (BNC HY6)

The meaning of the temporal expressions in (6) and (7) may be said to be non-compositional, since neither *minute* nor *hour* are used with their normal clock sense. The expression *at the minute* in (6) may be paraphrased *just now/at present*. It is

much more common in spoken than written English, with 65 of the 79 instances in the BNC being from the spoken demographic component, and a further 11 from fictional dialogue. The expression *at the hour* in (7) may be paraphrased *at the time* (see definition 4. a. of *hour* in the OED: “A definite time in general; an appointed time; an occasion”.) It is seldom used in spoken English apart from when the speaker is reciting the *Ave Maria*: i.e. *now and at the hour of our death*.

Points in time are regularly encoded in English by ‘*at x o’clock*’ expressions. ‘*O’clock*’ expressions always refer to a point in time. They may occasionally occur with *for*, as in ‘I’ve booked at the bistro down the road *for nine o’clock*’ (BNC CS4), but then *for* refers to a point in time at which some situation is due to occur rather than an interval. There are just three English nominals that regularly occur with both *at* coding a point (as indeed it always does) and *for* coding an interval. These are *time*, *instant* and *moment*. *Øyeblikk* in Norwegian can also be used to encode a point or an interval, as in examples (1) and (2). Unlike *moment*, however, when used to encode an interval it is not normally preceded by a preposition, the expression *et øyeblikk* corresponding to *for a moment* in English. The original English text in (2) testifies to the fact that it is perfectly possible to omit the preposition in English as well. When it comes to the point sense of *øyeblikk*, this is usually encoded in adverbial expressions by constructions containing the preposition *i* (‘in’), as in (8).

- (8) Jeg tror at allerede dagen før, *i det øyeblikk* han så altertavlen,
var hans beslutning klar. (JW1)

I think that his decision had been made the day before,
at the very moment he saw the altarpiece. (JW1T)

For the reader unfamiliar with Norwegian, the fact that one employs the equivalent of an English preposition that quintessentially denotes a relationship of containment to refer to a point rather than an interval may seem counter-intuitive. However, Norwegian does not have an *at* preposition. Rather it denotes contiguity by means of a handful of more specific prepositions, such as the *on* preposition *på*, the *by* preposition *ved*, and the *in* preposition *i*. *I* corresponds to English *at* in some 23% of the instances of spatial *at* predications in the ENPC, according to Egan and Rawoens (2017: 129).

3. Data and method

The data for the present study comprise all instances of *moment* in both original English texts and English translations, and all instances of *øyeblikk* in both original Norwegian texts and Norwegian translations in the ENPC. The corpus

contains extracts from 50 English and 50 Norwegian original texts, 30 fictional and 20 non-fictional, with their translations into the other language.² The four sub-corpora each contain between 600,000 and 700,000 words, with both translation sub-corpora containing more tokens than their source corpora. The fact that the four sub-corpora are of comparable length obviates the need for the linguist to employ normalised frequencies in comparing results.

Although the compilers of the ENPC took particular care to ensure that the original texts in the two languages were as similar as possible, this proved more difficult in the case of the non-fiction than the fiction texts. According to the compilers themselves, “There were special problems in finding non-fictional texts. Norwegian writers of non-fiction who want to reach an international audience often prefer to publish in English, rather than having their texts published first in Norwegian and then translated” (Johansson, Oksefjell & Ebeling 1999). As a result, there are bigger differences between the non-fiction text types in the two languages than the fiction texts. In particular the English corpus contains more texts with narrative components, such as histories and biographies. Since specifying the temporal location of a situation or its temporal extent is central to narrative discourse, as opposed to expository discourse, for example, one may expect this difference in text types to give rise to a difference in frequency in temporal predications in the two sets of original texts.

All instances of *moment* and *øyeblikk*, and its variant *augneblink*, of which there were only three tokens, were downloaded together with their translations, or in the case of translated text, their sources. These instances were classified by both authors according to whether they encoded points or intervals and whether they occurred in fixed expressions, such as *for a moment* or *a moment later*. The translations were categorised as syntactically congruent, as in (9), syntactically divergent, as in (10) or as zero, as in (11).

- (9) Det kom etter bare *et øyeblikk*. (FC1)
It came after only *an instant*. (FC1T)
- (10) *No moment* with her was ever dull. (AH1)
Han kjedet seg *aldri* sammen med henne.
(AH1T) [‘He was never bored with her.’]
- (11) Sarah waited *a moment* until she felt her own voice was under control. (MW1)
Sarah ventet til hun følte at hennes egen stemme var under kontroll.
(MW1T) [‘Sarah waited (Ø) until she felt...’]

2. For details of the corpus texts see <<http://www.hf.uio.no/ilos/english/services/omc/enpc/>>.

The translation in (9) is functionally congruent since the nominals *øyeblikk* and *instant* both occur in adverbials. The fact that they are both formally identical, insofar as they both occur in preposition phrases, is not decisive in determining congruence. The translation would also be classified as congruent if ‘after only an instant’ were to be replaced by ‘instantly’. The translation in (10) is said to be divergent, since a noun phrase, *no moment*, together with an adverbial *ever*, are translated by an adverbial *aldri* (‘never’). Finally, the translator in (11) has omitted altogether to translate *a moment*. The presence of a time interval is conveyed by the inclusion of an end-point encoded in the *til* (‘until’) clause, but the brevity signalled by *a moment* in the original is not directly conveyed.

4. *Moment* and *øyeblikk* in the source language texts in the ENPC

Table 1 contains details of the total number of tokens of *moment* and *øyeblikk* in the source texts in the two languages.

Table 1. Tokens of *moment* and *øyeblikk* in English and Norwegian originals

	Fiction	Non-fiction	Total
<i>moment</i> in English originals	164	52	216
<i>øyeblikk</i> in Norwegian originals	152	15	167

Perhaps the most striking feature of the results shown in Table 1 is the difference in the number of instances found in the non-fiction texts in the two languages. Examples (12) and (13), both from non-fiction English original texts, illustrate the coding of a point in time and a time interval respectively. As mentioned in Section 3, specifying the temporal location of a situation or its temporal extent is central to narrative discourse, and there are more texts with a narrative component among the English non-fiction originals.

- (12) With the prospect of distant victory in sight, many Allied soldiers yearned to remain alive to see *the moment*. (MH1)
 Med utsikten til en seier i det fjerne ønsket mange av de allierte soldatene å holde seg i live for å kunne oppleve *det øyeblikket*. (MH1T)
- (13) *For a moment*, I thought he was regretting the wound inflicted on grandfather André. (PM1)
Et øyeblikk trodde jeg at han var lei for at bestefar André var blitt bitt så stygt. (PM1T)

Table 2 contains details of the relative incidence of tokens denoting points and intervals in the two source sub-corpora. Tokens of both *moment* and *øyeblikk* are broadly dispersed among the two sets of original texts in the corpus. Of the 50 English original texts, 42 contain instances of *moment*. Of the 50 Norwegian original texts in the corpus, 33 contain instances of *øyeblikk*.

Table 2. *Moment* and *øyeblikk* coding point(s) or interval(s) in English and Norwegian originals, with horizontal percentages

	Point	Interval	Total
<i>moment</i> in English originals	122 / 56.5%	94 / 43.5%	216
<i>øyeblikk</i> in Norwegian originals	66 / 39.5%	101 / 60.5%	167

In 28 of the 42 English original texts with *moment*, we find it coding an interval, and in 41 coding a point. There are 27 texts that contain both senses. In the case of 12 of the 15 texts that contain just one sense, there are three or fewer tokens of the lexeme. In 29 of the 33 Norwegian original texts with *øyeblikk*, it codes an interval, in 28 it codes a *point*. There are 25 texts that contain both senses. In the case of seven of the nine texts that contain just one sense, there are three or fewer tokens of the lexeme.

It would appear from the numbers in Table 2 that, while *moment* and *øyeblikk* are both commonly used to code temporal intervals, the English lexeme is almost twice as likely to be employed to encode points in time as its Norwegian counterpart. The difference between the two lexemes when it comes to encoding points and interval is, as one might expect, statistically significant: Pearson’s chi.sq. with 1 df = 10.841, $p = 0.000993$. It is the contrast between the two lexemes with respect to the point sense we concentrate on in the remainder of this section.

We begin by pointing out, in Table 3, the syntactic functions in which we find the point sense of the two lexemes.³ The lexemes are employed in similar syntactic contexts in the two languages, in roughly similar proportions. This similarity is confirmed by statistical calculations (Pearson’s chi.sq. with 5 df = 3.508, $p = 0.62214$: if we leave out the *Other* row, which only contains small numbers, the probability returned by an exact test with 4 df is 0.51624).

In more than half of all cases both lexemes occur as part of an adverbial, as in (14) and (15).

3. We have not included a similar table for interval senses of the two lexemes, since these occur (almost) invariably in adverbials, signalling the time a situation lasted.

Table 3. Functions of point sense tokens of *moment* and *øyeblikk* with vertical percentages

	moment	øyeblikk
(Part of) Adverbial	68 / 55.7%	42 / 63.6%
(Part of) Subordinator	20 / 16.4%	10 / 15.2%
Object	13 / 10.7%	8 / 12.1%
Subject	11 / 9%	2 / 3%
Subject complement	7 / 5.7%	2 / 3%
Other	3 / 2.5%	2 / 3%
Total	122	66

- (14) The girl disappeared, and *a moment later* the other one came out, the frizzy one. (AT1)

Piken forsvant, og *et øyeblikk senere* kom den andre ut, den krushårete. (AT1T)

- (15) *Et øyeblikk etter* var hun tilbake med en pakke i hendene. (JW1) ['A moment afterwards...']

A moment later she returned with a package in her hands. (JW1T)

In addition to their occurrence in adverbials proper as in (14) and (15), we find both lexemes fulfilling what we might call a subordinating function in an adverbial clause. There are twice as many tokens of *moment* in this role than of *øyeblikk*. The English the noun phrase *the moment* is used like this in (16). There are nine such tokens.

- (16) *The moment* the student spoke, he would be recognised by his voice. (OS1)
Så snart studenten begynte å snakke, pleide han å bli gjenkjent på stemmen. (OS1T) ['As soon as ...']

In (16) *the moment* fulfils the same syntactic functions as a subordinator like *when*. There are no tokens in the corpus of original Norwegian texts where nominal *øyeblikk* is used like this. This sort of subordinating function is more commonly executed in Norwegian by preposition phrases containing *øyeblikk*, as in (17) and (18).

- (17) Men *i det øyeblikket* det akkurat var over, fantes ingen sorger mere. (HW1)
 But *at the exact moment* it was over, she had no more worries. (HW1T)

- (18) *Fra det øyeblikket* Osmundsen hadde uttalt sin mistanke til ham, hadde den, enda så uberettiget den var, preget, for ikke å si dirigert hele hans væremåte. (HK1)

From the moment Osmundsen had expressed his suspicion of him, it had, however unjustified, coloured – if not controlled – his entire mode of existence. (HK1T)

There are seven tokens of *i det øyeblikk(et)* as in (17) and just two of *fra det øyeblikk(et)* as in (18) functioning as subordinators in the original Norwegian texts. In the English texts we find five tokens of *from the moment* and one of *at the moment* serving the same function.

The two lexemes occur regularly in objects in both languages and normally give rise to syntactically and lexically corresponding translations, as in (19) and (20).

- (19) But the metamorphosis was real enough, and she enjoyed
every moment of it. (MD1)

Men metamorfosen var ekte nok, og hun nøt hvert eneste øyeblikk. (MD1T)

- (20) Kjersti sliter ræva av seg med å fange øyeblikket. (CL1)

Kersti breaks her ass to capture the moment. (CL1T)

Moment is much more common as subject, as in (21), than is *øyeblikk*, but the fact that the latter can also fulfil this function is attested to by (22) as well as the translation of (21).

- (21) Macon wondered if even *this moment* would become, one day,
something he looked back upon wistfully. (AT1)

Macon lurte på om *dette øyeblikket* også en dag ville bli noe han
lengselsfullt så tilbake på. (AT1T)

- (22) Nå etterpå, etter alt sammen er det som om akkurat *det øyeblikket*
har spikra seg fast i bevisstheten mi.

(JM1) ['...just that moment has nailed itself firmly in ...']

Now, afterwards, after everything that's happened, it's *that*
particular *moment* I find stuck in my mind. (JM1T)

Like subjects, subject complements, as in (23), are more common in the English data, although they also occur in Norwegian, as in (24).

- (23) This was a delicate *moment*. (DL2)

Dette var et vanskelig *øyeblikk*. (DL2T)

- (24) Det var et ubeskrivelig rørende *øyeblikk* da dronning Maud med
stor forsiktighet bar sitt lille barnebarn frem til døpefonten. (TG1)

It was an indescribably touching *moment* when Queen Maud
carefully bore her little grandchild up to the baptismal font. (TG1T)

Finally, there are five tokens in all, three in English and two in Norwegian that do not play any of these syntactic roles. In all three English examples, one of which is cited as (25), *moment* occurs in a post-modifier in a noun phrase. There is one such example in Norwegian. In the other Norwegian example, cited as (26), *øyeblikk* occurs in a sentence fragment.

- (25) All that still has to be decided is the timing of *that moment*. (ABR1)
 Det eneste som står igjen å bestemme, er når *dette øyeblikket* skal inntreffe.
 (ABR1T) ['...is when this moment is to occur.']
- (26) Et underlig *øyeblikk*: de to på et hotellrom. (KH1)
 A strange *moment*, the two of them in a hotel room. (KH1T)

To sum up this overview of the point sense of *moment* and *øyeblikk*, we have seen that the two lexemes occur in the same syntactic contexts. The single exception of note was the use of the noun phrase *the moment* as a subordinator in adverbial clauses. In the next section we will examine the translation correspondences of the two lexemes, and in particular whether translators tend to favour congruent or divergent translations.

5. Translation correspondences of *moment* and *øyeblikk*

Table 4 contains details of the most common translations of both senses of *moment* in the target texts. We see that *moment* is translated three times or more by as many as eight different Norwegian words/phrases. In many cases it would be more accurate to say that it is the phrase containing *moment* that is so translated. Thus the Norwegian phrase *når som helst* translates the English phrase *at any moment* and not just the word *moment*.

Table 4. Translations of *moment* in ENPC with three or more tokens

Norwegian translations	English glosses	Totals	
<i>øyeblikk</i>	'wink of an eye'	116	53.7%
<i>stund</i>	'while'	20	9.3%
<i>tidspunkt</i>	'point in time'	5	2.3%
<i>akkurat da/nå</i>	'just then/now'	4	1.9%
<i>når som helst</i>	'at any time'	4	1.9%
<i>litt</i>	'a little'	3	1.4%
<i>foreløpig</i>	'for the time being'	3	1.4%
<i>så snart</i>	'as soon as'	3	1.4%
Other		40	18.5%
Ø	Ø	18	8.3%
Total		216	

We have already seen that *øyeblikk*, the most common correspondence of *moment*, is used to translate both the point sense, as in (1) and (12), and the interval sense, as in (2) and (13), of the latter. There are in all 63 instances of *øyeblikk* translating

the point sense, and 53 instances of it translating the interval sense, indicating that the two lexemes share a common polysemy with respect to these two senses. The Norwegian noun *stund* is also used to translate both interval senses of *moment*, as in (27), and point senses, as in (28). It differs in this respect from its closest English counterpart *while*, which is only used to refer to an interval in Present-day English.

- (27) She regarded me *for a long moment* before finally opening her desk drawer.
(TH1)

Hun så på meg *en lang stund* før hun til slutt åpnet skrivebordskuffen.
(TH1T) ['...a long while...']

- (28) There, *from the moment* he arrived, he was an object of suspicion and sneaky ridicule.
(JH1)

Der ble han *fra første stund* gjort til gjenstand for mistro og lumske latterliggjøring.
(JH1T) ['...from first while...']

There are 15 examples in the corpus of *stund* being used to translate the interval sense as opposed to just five of the point sense. All the other words or phrases in Table 4 are just used to translate one sense or the other, with *litt* ('a little') and *foreløpig* ('for the time being') being restricted to the interval sense, and *tidspunkt* ('point in time'), *akkurat da/nå* ('just then/now'), *når som helst* ('at any time') and *så snart* ('as soon as') restricted to the point sense.

There are 40 examples listed as 'other' in Table 4. These comprise as many as 32 types of word or phrase, with 31 examples translating a point and just nine an interval. There is no space here to present the various translations in detail, so (29) and (30) must serve as illustrations of some other options chosen by translators to translate points and intervals respectively.

- (29) Well, I've thought over *every moment* of that party time and time again.
(JSM1)

Nei, jeg har tenkt gjennom *hvert minutt* av den grillfesten om og om igjen.
(JSM1T) ['...every minute ...']

- (30) *For a moment* the rabbit did not move at all.
(SK1)

Et par sekunder rørte den seg ikke i det hele tatt.
(SK1T) ['A couple of seconds...']

There are 18 instances of zero translations of *moment* into Norwegian, ten of which omit to translate intervals, as in (31), the other eight points, as in (32).

- (31) Sarah waited *a moment* until she felt her own voice was under control.
(MV1)

Sarah ventet (Ø) til hun følte at hennes egen stemme var under kontroll.
(MV1T)

- (32) She might have pressed for more information, but the television screen flashed with snow *at that moment* and then flipped over to a closeup of John Daggett. (SG1)
- Hun kunne ha presset meg for flere opplysninger men det glimtet til med snø på fjernsynsskjermen (Ø), og i neste nå kom et nærbilde av John Daggett. (SG1T)

All 18 untranslated instances of *moment* are adverbials like (31) and (32). They are spread across 15 translators, indicating that no one translator is especially prone to employing the omission option. The fact that the event predicated takes place at a point, or over a short interval of time, is generally obvious from the context. (31) is a particularly interesting example since the remainder of the translated sentence consists of a word for word translation, apart from the addition of the subordinator *at* ('that'). This would seem to indicate a wish to remain as faithful as possible to the original text, and yet the translator has chosen to omit the adverbial *a moment*, presumably reasoning that the readers' own world knowledge will allow them to infer that the time it takes to recover control of one's voice is generally brief. In (32) the presence of the contrastive/adversative coordinator *but* entails that the possibility of eliciting more information was interrupted by the flashing on the television. At the risk of stating the obvious, the interruption must have taken place at the moment of interruption, so nothing of communicative import is lost by omitting the 'time when' adverbial. In some cases the translator compensates for the omission of the adverbial by indicating the existence of the point/interval in some other way. Example (33) is a case in point.

- (33) Willie stood stunned *for a moment* for he had never been praised by anyone ever. (MM1)
- Willie ble stående litt overrumplet, for han hadde aldri fått ros av noen. (MM1T) ['... continued standing ...']

In (33) the translation of the simple English verb phrase *stood* by the continuous form *ble stående* serves to denote that the situation in question lasted for a certain time, thus compensating for the omission of the interval-denoting *for a moment*. When we now turn to English translations of *øyeblikk*, we see that Table 5 contains fewer correspondences than Table 4.

Table 5 shows that only five translation options into English are employed three or more times compared to the twice as many into Norwegian shown in Table 4. Apart from *moment*, which is used in three quarters of all instances, there are only two lexemes listed in the table, *instant*, illustrated here by (34) and *momentarily*, illustrated by (35).

Table 5. Translations of *øyeblikk* in ENPC with three or more tokens

English translations	Totals	
<i>moment</i>	126	75.4%
<i>instant</i>	6	3.6%
<i>momentarily</i>	3	1.8%
Other	14	8.4%
Ø	18	10.8%
Total	167	

- (34) Herman lener hodet bakover og ser opp på faren, og *et øyeblikk* er det nesten som da han stod under treet i Frognerparken. (LSC1)

Herman leans his head back and looks up at Father, and *for an instant* it's almost like when he stood under the tree in Frogner Park. (LSC1T)

- (35) Selv fuglene i dalen sluttet å synge da biskopen *et øyeblikk* stoppet opp og rømnet seg. (JW1) ['...a moment...']

Even the birds in the valley had stopped singing when the Bishop paused *momentarily* and cleared his throat. (JW1T)

Instant, like *moment*, is used to translate both intervals, as in (34), and points. *Momentarily* is only used to translate intervals, which is as one would expect if the translator adheres to a British English norm. A minority of the translations in the ENPC are into American English, but these contain no examples of the American use of *momentarily* to mean 'at any moment'.

There are 14 examples listed as 'other' in Table 5. Of the 14, four options are used twice; *minute*, *second*, *just about to* and *immediately*. Of the six options taken just once, two contain a derivation of *moment*, as in (36). Eight of the 14 translate a point, the remaining six an interval. (37) is one of a small handful of cases in the corpus where an interval is translated as a point.

- (36) Hun fikk ikke noe svar nå heller, men *i et lite øyeblikk* klarte hun å bli i villrede om det var hun selv eller speilbildet som hadde stilt spørsmålet. (JG1) ['...in a little moment...']

She received no response to this either, but felt a *momentary* confusion as to whether it was she or her reflection who had asked the question.

- (37) Vognen ble *på et øyeblikk* fylt med gale borgerkvinner som skulle finne seg et sjarmerende fosterbarn. (CL1) ['...in a moment...']

Immediately the carriage was filled with crazy bourgeois women who wanted to find themselves a charming foster child. (CL1T)

In (37) the phrase *på et øyeblikk* tells the reader that it only took a very short time for the carriage to fill up. The English translation, on the other hand, tells us when the filling up happened. The preceding sentence reads ‘The train stopped at every station’. In other words, whenever the train stopped it was filled up at once, which implies that the interval between the train stopping and the women boarding it was very short.

There are 18 instances of zero translations of *øyeblikk* into English, the same number as translations of *moment* into Norwegian. Eight of these omit to translate intervals, as in (38), the other ten points, as in (39).

- (38) Spot kikket ut av vinduet et øyeblikk, med mørkt blikk som om han skammet seg.

(EFH1) [‘... *peered out of the window a moment* ...’]

Spot *glanced out the window*, his eyes darkening as if he were ashamed.

(EFH1T)

- (39) Men han tenkte ikke på det i det øyeblikk han så at Dina sto i døren.

(HW2) [‘... *in the moment he saw*...’]

But he was not conscious of his appearance *when he saw* Dina standing in the doorway.

(HW2T)

In the translation of (38) the omission of the short interval adverbial is compensated for by the choice of the verb *glance* which, according to the OED (sense 5), means ‘Of the eye: To move quickly, to cast a momentary look’, to translate *kikke* which does not necessarily imply a brief look (one of the examples in the official Norwegian dictionary (*Bokmålsordboka*), published by the Language Council of Norway, is *kikke lenge på noe* (‘look for a long time at something’)). As for (39), one could argue that not much of import is lost by omitting *øyeblikk* and just using *when* to connect the subordinate to the main clause. There is, nevertheless, more of an implication that he would be more conscious of it on other occasions in the Norwegian original than the English translation.

Finally, if we compare Tables 4 and 5 we see that *øyeblikk* is more likely to be translated by *moment* than the other way round. In other words, the correspondence between the two lexemes is asymmetric. Ebeling and Ebeling (2013: 27) discuss asymmetric correspondences and suggest that the item with the lower translation bias (*moment* in our case) is likely to exhibit a greater degree of polysemy than its counterpart. Be that as it may, both correspondences are over 50%, and their mutual correspondence, arrived at by adding the number of corresponding translations in the two texts, multiplying by 100 and then dividing this figure by the sum of the total number of occurrences of the two items in the source texts, is 63% (see Altenberg 1999). The figures are given in Table 6.

Table 6. Mutual correspondence (cf. Altenberg 1999) of *moment* and *øyeblikk*

	Originals	Corresponding
<i>moment</i>	216	116
<i>øyeblikk</i>	167	126
Total	383	242 (mc = 63%)

6. When *moment* is not *øyeblikk* and *øyeblikk* is not *moment*

Although *moment* and *øyeblikk* have been shown as translation correspondents in the majority of cases, their distribution in original and translated texts differs. Both lexemes are found with more tokens in translations than in original texts, as shown in Table 7. Here, we see the total number of occurrences of *moment* in English original and translated texts and of *øyeblikk* in Norwegian original and translated texts.

Table 7. *Moment* and *øyeblikk* in fictional and non-fictional source and target texts

	Fiction	Non-fiction	Total
<i>moment</i> in English originals	164	52	216
<i>moment</i> in English translations	222	29	251
<i>øyeblikk</i> in Norwegian originals	152	15	167
<i>øyeblikk</i> in Norwegian translations	157	47	204

There are more tokens of *moment* in the translated English texts than in the original English texts and more tokens of *øyeblikk* in the translated Norwegian texts than in the original Norwegian texts. We can also see that the tokens of *moment* in English translations outnumber the tokens of *øyeblikk* in Norwegian original texts, the source texts for the English translations. Table 4 shows that close to half the occurrences of *moment* in original English texts are translated with something other than *øyeblikk*, or indeed left out completely. The number of occurrences of *øyeblikk* in Norwegian originals and *moment* in English translations also indicates that *moment* in the English translated texts must have other sources than *øyeblikk*.

Zero translation, or omissions, of both *moment* and *øyeblikk* were discussed in Section 4. We do, however, also find several instances of adding of both *moment* and *øyeblikk* in the material from the ENPC: expressions with *moment* or *øyeblikk* in the translated texts that do not have an obvious source in the original text. In the English translations, we find 20 examples of the addition of *moment*, spread

across 11 texts. One translator (of EG1T, EG2T and KB1T) is responsible for nine additions, and another translator (of KA1T) adds *moment* three times. In most instances, the immediate context does not contain an expression or contextual trigger for the use of *moment* in the translation, for example in (40) and (41). However, a few additions of *moment* can be explained from context in the original Norwegian text. One example is the use of the past perfect in (42), implying the end of an interval.

- (40) Han må teie før han går vidare. (EH1) ... *hold his tongue* (Ø) *before* ...
He is silent *for a moment* before he continues. (EH1T)
- (41) Han sa ikke bare det som var opportunt, langt derifra. (TG1)
[‘... whatever was opportune (Ø), far from it.’]
He did not say what was opportune *at the moment*, far from it. (TG1T)
- (42) Osmundsen hadde sittet helt ubevegelig, liksom fastfrosset. (KA1)
[‘... had sat motionless (Ø) ...’]
Osmundsen sat quite motionless *for a moment*, as though frozen stiff. (KA1T)

Although Table 7 shows that the tokens of *øyeblikk* in the Norwegian translations do not outnumber the tokens of *moment* in the English original source texts, there are 16 additions of *øyeblikk* in the Norwegian translations. As with the English additions of *moment*, most of the additions of *øyeblikk* do not have an obvious source in the context, as we can see in (43) below. In a few instances, there are contextual triggers, such as “alternating current”, indicating sudden shifts, in (44). For the additions, both of *moment* in the English and *øyeblikk* in the Norwegian translations, there is an even distribution between the point and interval meaning.

- (43) Thunder growled above the house and lightning cracked. (BO1)
Torden rullet over huset, og *i neste øyeblikk* lynte det. (BO1T)
[‘... *and in the next moment* lightning cracked.’]
- (44) Pete’s natural state of mind was an alternating current of elated certainty and angry disappointment. (JSM1)
Pete kunne slå om *på et øyeblikk* fra opprømt forvisning til rasende skuffelse. (JSM1T) [‘Pete could turn *in a moment* from ...’]

In Tables 4 and 5, expressions involving the lexeme *øyeblikk* account for more than 50% of the translations of *moment* and expressions involving the lexeme *moment* account for more than 75% of the translations of *øyeblikk*. However, we also note in Section 4 that there are other translation options for both lexemes. By taking a closer look at some of the multi-word combinations including *moment* and *øyeblikk*, we

explore further the range of translation solutions which do not include the corresponding lexeme (*øyeblikk* and *moment*, respectively). The examples in Tables 8 and 9 were selected from expressions with more than 3 tokens in both original and translated texts, where both sources and translations contained correspondences which did not include *øyeblikk* in Table 8 and *moment* in Table 9. Table 8 presents multi-word combinations including the lexeme *moment*. The left-hand column contains the sources of the English multi-word combination (from Norwegian original texts) and the right-hand column shows what the same English multi-word combination is translated into in Norwegian; both columns are sorted alphabetically.

Table 8. Norwegian correspondences of multi-word combinations with *moment* in ENPC, excluding the lexeme *øyeblikk* (and *augneblink*)

Norwegian original texts		Norwegian translated texts
<i>en stund</i> <i>lite grann</i> <i>litt</i> aspect (<i>ble stående</i>) Ø	for a moment	<i>en stund</i> <i>et par sekunder</i> <i>umiddelbart</i> Ø
<i>da</i> <i>den stunden</i> <i>i det samme</i> <i>med det samme</i> <i>nå</i> <i>samtidig</i> Ø	at that moment	<i>akkurat da/akkurat nå</i> <i>nettopp</i> <i>på rede hånd</i> <i>på det tidspunkt</i> Ø
<i>hele tiden</i> <i>med en gang</i> <i>når som helst</i> <i>rett som det var</i>	at any moment	<i>når som helst</i>

The most frequent 3-gram including *moment* in both the English original texts and the English translated texts is *for a moment*, which occurs 41 times in English originals and 82 times in English translations. This over-representation in translation is also noted by Ebeling, Ebeling and Hassegård in their investigation of recurrent word-combinations in the ENPC (2013: 193). Unsurprisingly, the most common Norwegian sources of *for a moment* also involve the lexeme *øyeblikk* or *augneblink*: 57 of the sources are *et øyeblikk*/ *ein augneblink* or *noen øyeblikk* ('some moments'). In the other direction, the most common Norwegian translations of *for a moment* also involve the lexeme *øyeblikk*: 26 of the translations are *et øyeblikk*, *et lite øyeblikk* ('a small moment') or *et eneste øyeblikk* ('a single moment').

As we saw in Table 4, *stund* ('while') is the second most common translation of *moment* overall (excluding 'other'). When we look at the overview for *for a*

moment in Table 8, *en stund* occurs as both source and translation, seven times in the Norwegian source texts, here exemplified by (45), and seven times in the Norwegian translations, here exemplified by (46).

- (45) Han tenkte seg om *en stund*. (EFH1)
He thought *for a moment*. (EFH1T)

- (46) It dominated the room so that he and Alice Mair stood
for a moment, silently regarding it. (PDJ3)

Det dominerte rommet, slik at han og Alice Mair ble stående *en stund* og se på det i taushet. (TG1T) ['... Alice Mair remained standing **a while** ...']

Apart from zero translation, the other correspondences occur only either as a source or as a translation, not as both. While they are so infrequent that it is impossible to suggest a reason for this, translator idiosyncrasies may well play a role, for example in (47), where *for a moment* is translated with *umiddelbart* ('immediately'):

- (47) But *for a moment*, gladdened by her easy assumption that they were a couple ... (PDJ3)

Men han ble *umiddelbart* glad for at hun anså dem for et par ...
(PDJ3T) ['... But he was **immediately** happy that she considered them ...']

In (46), as in (33), the interval sense of *for a moment* is emphasized through the use of aspect (*ble stående*) in the Norwegian translation. We also find aspect (*ble stående*) in the Norwegian original as a potential source of *for a moment* in (48).

- (48) Han la på røret, *ble stående*, tenkte: Hvis signalementet ikke var en bløff, hvorfor arresterte de meg så ikke? (KA1) ['... *remained standing* ...']

He put down the receiver and *stood for a moment*, thinking: If the description wasn't a bluff, why don't they go ahead and arrest me? (KA1T)

For *at that moment* in English translations, the Norwegian source is *i det samme* in three instances (49), *da* ('then') is the source in two (50), while all the other Norwegian sources are found only once. Apart from zero translation, all the correspondences occur only in one direction, as either source or translation of *at the moment*. When *at that moment* is translated from English into Norwegian, no Norwegian translation is found more than once in the ENPC.

- (49) *I det samme* satte toget seg i bevegelse, (EFH1)
At that moment the train started moving (EFH1T)

- (50) - Jeg tapte ikke gullet, jeg vant sølvet, sa Aamodt – *da*
lykkelig uvitende om ... (KB1)

"I didn't lose the gold, I won the silver, Aamodt declared –
at that moment blissfully unaware that ... (KB1T)

When we take a closer look at *at any moment*, the situation is more similar to that of *for the moment*, with one correspondence, *når som helst* ('at any time'), not only the most frequent overall, but also the only one to be found in both directions, both as the Norwegian source of the English multi-word combination, as in (51), and as the Norwegian translation, as in (52).

- (51) Den kommer *når som helst*! (MN1)
It'll be here *at any moment*! (MN1T)
- (52) It was no good to invest hard work in land on which the Forestry
Department might *at any moment* plant pines. (LT1)
Det var ingen mening i å investere i hardt arbeid på jordstykker som
skogdepartementet *når som helst* kunne komme og plante furuer på. (LT1T)

In fact, *når som helst* is the only translation of *at any moment* found in the material, but the English phrase only occurs three times in the English source texts. Overall, there is a greater spread of different Norwegian sources for English multi-word expressions with *moment* than there is in Norwegian translations of expressions with *moment*, suggesting that there could be a tendency to use a more limited range of expressions in translation and overuse correspondences with the most salient lexeme (in our case, *øyeblikk*). This may be an indication of what Gellerstam calls "Pavlovian translation", that a lexeme which in many cases, but not always, corresponds to the source lexeme "remains the standard choice when translating from English, partly due to the power of habitual use, partly due to the translator failing even to consider other alternatives" (2005: 206). Johansson (2007b) found a similar pattern in his study of the English time expression *spend (time)* and the Swedish and Norwegian counterparts *tillbringa (tid)* and *tilbringe (tid)*, where the Swedish and Norwegian expressions are used far more frequently in translations than in original texts, despite a wide range of alternative translations for the English *spend*.

This tendency is also clear when we turn to the English sources and translations of Norwegian multi-word expressions containing *øyeblikk* in Table 9. The left-hand column contains the English sources of the Norwegian multi-word combination and the right-hand column shows what the Norwegian multi-word combination is translated into in English, sorted alphabetically in both columns.

We have already noted that *øyeblikk* is less frequent than *moment* in both original and translated texts. There are also fewer multi-word combinations including *øyeblikk*, perhaps because the nearest correspondence to the overwhelmingly frequent *for a moment* in the English texts is not a multi-word combination in Norwegian, but simply the 2-gram *et øyeblikk* ('a moment').

The most frequent multi-word combination in the Norwegian material is '*et + (adj) + øyeblikk*' ('a + (adj) + moment'), both for the Norwegian original texts (eight occurrences) and Norwegian translations (twelve occurrences). When we disregard the correspondences which include the lexeme *moment*, we see that there are

Table 9. English correspondences of multi-word combinations with *øyeblikk* in ENPC, excluding the lexeme *moment*

English original texts		English translated texts
<i>briefly</i> <i>instant</i> <i>minute</i> <i>momentarily</i>	et (adj) <i>øyeblikk</i>	<i>one (adj) instant</i>
<i>in no time</i> <i>record time</i> Ø	på et <i>øyeblikk</i> / på <i>øyeblikket</i>	<i>immediately</i> Ø
<i>currently</i> <i>for now</i> <i>now</i> Ø	i <i>øyeblikket</i>	<i>just then</i>

four different English sources for ‘*et + (adj) + øyeblikk*’, but the multi-word combination is only translated once as something other than *moment*, as we see in (53).

- (53) Han hadde fornemmet at alt sto stille *et kort øyeblikk*, før murveggene begynte å skjelve og rase sammen. (JW1) [‘... *a short moment* ...’]
He noticed that everything stood still for *one brief instant*, before the walls started to shake and crumble. (JW1T)

As for the Norwegian multi-word expressions *på et øyeblikk* (‘on a moment’) and *på øyeblikket* (‘on the moment’), these occur in Norwegian originals three times each in the indefinite form and once in the definite form, while the indefinite form (*på et øyeblikk*) appears three times in the translated Norwegian texts. In the translations into English, *moment* is only found once, and it does not appear as a source for any of the Norwegian translations. In the case of ‘*et + (adj) + øyeblikk*’, *instant* appears as both source and translation. For *på et øyeblikk* / *på øyeblikket*, there are no correspondences which appear as both English sources and English translations. *På et øyeblikk* has been added once, as we saw in example (44), and it is also found once as the translation for *in record time* and once for *in an instant*. From Norwegian to English, *på et øyeblikk* / *på øyeblikket* has been translated twice as *immediately*, one example is found in (37) above. Zero translation also only occurs once, in (54), and there is only one instance of *moment* as the translation correspondence.

- (54) Fjøsset ligger bare vel to meter fra seterhuset og med dørene slik at budeia *på et øyeblikk* kunne smette over i fjøsset fra selsgangen. (AOH1) [‘... the dairymaid **in a moment** could ...’]
The cowshed stands a few yards away with its doors such that the dairymaid (Ø) could quickly run in from the hallway. (JW1T)

The correspondences of the Norwegian *i øyeblikket* yet again show a greater diversity in the English sources than in the translations into English. In the eight occurrences of *i øyeblikket* in Norwegian translations, an expression with *moment* is the source in only two instances and the expression has been added with no obvious source once. The remaining correspondences all occur once each, here exemplified by *currently* in (55). When we look at the Norwegian original texts, *i øyeblikket* occurs only three times, twice translated by *moment* and in the last case from *akkurat i øyeblikket* ('exactly in the moment') to *just then*, as we see in (56).

(55) ... development as it is *currently* practiced. (LT1)

... utviklingsarbeidet slik det drives *i øyeblikket*. (LT1T)

(56) Det siste kom som om han *akkurat i øyeblikket* oppdaget at det var den eneste løsningen. (HW2)

The final words came out as if he had *just then* discovered that this was the only solution. (HW2T)

While the numbers are too low to draw any definite conclusions, it would seem that our investigation of translation correspondences of multi-word combinations with *moment* and *øyeblikk* in the ENPC supports Gellerstam's (2005) and Johansson's (2007b) suggestion that there is a tendency to overuse the lexical item which appears to be the most obvious (or prototypical) correspondence. We may be observing what Halverson (2004) calls a "gravitational pull" towards a more limited selection of translation correspondences, even when there are clearly, as we have seen in the original text material, a wider range of options available.

7. Summary and conclusion

In this chapter, we have examined the two lexemes *moment* and *øyeblikk* in the ENPC. The tokens of both lexemes were categorised in Section 2 according to the notions of time as a point and time as an interval or limited period. Other temporal expressions are also compared to *øyeblikk* and *moment* to shed light on their use in English and Norwegian multi-word constructions denoting time. In Section 4, we showed that the temporal interpretation of 'point' or 'interval' is influenced by the different constructions in which *moment* and *øyeblikk* appear. Different multi-word constructions as well as different syntactic functions were highlighted. Despite some minor differences, the two lexemes are similar in terms of syntactic functions, and they frequently give rise to congruent translations.

By using both original and translated texts from the ENPC, we show that the translations of *moment* and *øyeblikk* are most frequently, and perhaps unsurprisingly, *øyeblikk* and *moment*, although the correspondence is asymmetric in that

the Norwegian *øyeblikk* is more likely to be translated by *moment* than the other way around. For *moment*, just over half of the tokens are translated by *øyeblikk* or expressions with *øyeblikk*. The second most common Norwegian translation correspondent is *stund*, which can also be used with both the point sense and the interval sense. In our material, however, only a quarter of the examples of *stund* in Norwegian translated texts correspond to *moment* in the point sense in the original English texts.

Our study also uncovered several examples of omission or addition of both *moment* and *øyeblikk* in translation. In the translated texts, there is zero translation of *moment* 18 times and zero translation of *øyeblikk* 18 times. It seems that, in these cases, no essential information is lost by the omission, and the information about point or interval duration is recoverable from context or indicated in other ways. In one example, we also see that the Norwegian translator uses aspect in the verb phrase (the continuous *ble stående*) in the translated text to indicate an interval temporal interpretation. Our material also contains several instances of addition of expressions with *moment* or *øyeblikk* in the English and Norwegian translated texts, where there is no obvious source in the original Norwegian or English text. There are 20 additions of *moment* in the English translated texts and 16 additions of *øyeblikk* in the Norwegian translated texts. A few of these additions may have their source in contextual clues in the source text, and in at least one instance, the use of aspect in the Norwegian verb phrase (again the continuous *ble stående*) is likely to have triggered the use of *moment* in the English translation.

A close look at sources for and translations of word-combinations containing *moment* and *øyeblikk* gives us a clear view of the asymmetrical relationship between the two lexemes. As shown in Tables 4 and 5, the tendency for *øyeblikk* to be translated into English as *moment* is stronger than the tendency for *moment* to be translated into Norwegian as *øyeblikk*. In Tables 8 and 9 we see that the range of translation solutions is in many cases much more restricted for the Norwegian translations of *moment* than for the English translations of *øyeblikk*. For both languages, there is also a wider range of source correspondents than translation correspondents for many of the expressions containing *moment* or *øyeblikk*. Our findings show that, although the two lexemes display a high degree of mutual correspondence, they both tend to be overused in translation compared to their incidence in original language texts.

Acknowledgement

We would like to thank two anonymous reviewers for their helpful and constructive comments on our chapter.

References

- Altenberg, Bengt. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In *Out of Corpora: Studies in Honour of Stig Johansson*, Hilde Hasselgård & Signe Oksefjell (eds), 249–268. Amsterdam: Rodopi.
- Bokmålsordboka | Nynorskordboka*. <<https://ordbok.uib.no>> (1 April 2021).
- British National Corpus*. 2001. Oxford: Oxford University Computing Services.
- Ebeling, Jarle & Ebeling, Signe Oksefjell. 2013. *Patterns in Contrast* [Studies in Corpus Linguistics 58]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.58>
- Ebeling, Jarle, Ebeling, Signe Oksefjell & Hasselgård, Hilde. 2013. Using recurrent word-combinations to explore cross-linguistic differences. In *Advances in Corpus-based Contrastive Linguistics: Studies in Honour of Stig Johansson* [Studies in Corpus Linguistics 54], Karin Aijmer & Bengt Altenberg (eds), 177–199. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.54.11ebe>
- Egan, Thomas & Rawoens, Gudrun. 2017. Locative *at* seen through its Swedish and Norwegian equivalents. In *Cross-linguistic Correspondences. From Lexis to Genre* [Studies in Language Companion Series 191], Thomas Egan & Hildegunn Dirdal (eds), 121–145. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.191.05ega>
- The English–Norwegian Parallel Corpus*. <<http://www.hf.uio.no/ilos/english/services/omc>> (1 April 2021).
- Evans, Vyvyan. 2004. *The Structure of Time. Language Meaning and Temporal Cognition* [Human Cognitive Processing 12]. Amsterdam: John Benjamins. <https://doi.org/10.1075/hcp.12>
- Gellerstam, Martin. 2005. Fingerprints in translation. In *In and Out of English: For Better, For Worse*, Gunilla Anderman & Margaret Rogers (eds), 201–213. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781853597893-016>
- Haspelmath, Martin. 1997. *From Space to Time: Temporal Adverbials in the World's Languages*. Munich: Lincom.
- Halverson, Sandra. 2004. The cognitive basis of translation universals. *Target* 15(2): 197–241. <https://doi.org/10.1075/target.15.2.02hal>
- Johansson, Stig. 2002. Towards a multilingual corpus for contrastive analysis and translation studies. In *Parallel Corpora, Parallel Worlds. Selected Papers from a Symposium on Parallel and Comparable Corpora at Uppsala University, Sweden, 22–23 April, 1999*, Lars Borin (ed.), 47–59. Amsterdam: Rodopi. https://doi.org/10.1163/9789004334298_003
- Johansson, Stig. 2007a. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies* [Studies in Corpus Linguistics 26]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.26>
- Johansson, Stig. 2007b. Spending time in English and Swedish. *Nordic Journal of English Studies* 6(1). <<http://ojs.ub.gu.se/ojs/index.php/njes/article/view/15/17>> (1 April 2021). <https://doi.org/10.35360/njes.6>
- Johansson, Stig, Oksefjell, Signe & Ebeling, Jarle. 1999. English-Norwegian Parallel Corpus: Manual. <<https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/ENPCmanual.pdf>> (1 April 2021).
- Malinowski, Bronislaw. 1927. Lunar and seasonal calendar in the Trobriands. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland* 57: 203–215. <https://doi.org/10.2307/2843682>

- Munn, Nancy D. 1992. The cultural anthropology of time: A critical essay. *Annual Review of Anthropology* 21: 93–123. <https://doi.org/10.1146/annurev.an.21.100192.000521>
- OED. 1994. *The Oxford English dictionary*. On compact disc. Oxford: OUP.
- Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca NY: Cornell University Press.
<https://doi.org/10.7591/9781501743726>

Then and now in English and French

Parallel patterns?

Diana M. Lewis

Aix Marseille University, Laboratoire Parole et Langage

English *now* and *then* and French *maintenant* and *alors* all exemplify the cross-linguistic tendency for temporal adverbs to grammaticalize into markers of rhetorical relations. This paper analyses the polysemy of these adverbs in a comparable corpus of written-to-be-spoken speeches and unscripted spoken interviews. For *now* and *maintenant*, while the discourse patterns and the direction of change are remarkably similar, French and English seem to be at different points in the grammaticalization cycle, with *maintenant* being less grammaticalized than *now*. In the case of *alors* and *then*, it is the French lexeme that appears to be more bleached, occurring in a wider range of rhetorical contexts and in different discourse patterns from English *then*.

1. Introduction

1.1 Time in language

It is well known that cross-linguistically, time expressions develop argumentational and text-organizational functions. Typically, time expressions can grammaticalize to acquire meanings such as adversative, causative, concessive and conditional (Heine & Kuteva 2002: 291ff), although different languages may do so in different ways.

This chapter compares two pairs of temporal adverbs across French (of France) and English (of the UK), *now* / *maintenant* and *then* / *alors* in two related genres of political discourse: speeches and discussion. These adverbs have all evolved the types of cohesion-marking functions mentioned above. *Now* and *maintenant* have acquired contrastive and topic-shifting functions, while *then* and *alors* have both been recruited to express inference, marking conditional apodosis and inferred result, and as discourse management tools.

1.2 Overview

The next section sets the study within the domain of ‘discourse phraseology’, an expansion of the usual sense of phraseology to include regular patterns at the level of discourse, where schematic and sparsely-filled patterns are found. Section 3 then presents the comparable corpora of political discourse used in the study. This is followed by the findings of the comparative studies of English *now* and its French approximate equivalent *maintenant* in Section 4, and of English *then* and French *alors* in Section 5. Section 6 provides a summary and conclusion.

2. Contrastive phraseology

2.1 Discourse phraseology

Language usage is at the heart of phraseology, which deals with conventional linguistic production rather than rule-governed production. ‘Rule-governed production’ refers here, in the context of discourse, to multi-word linguistic sequences that are transparent insofar as they are built up by the speaker from smaller units in accordance with the permissions and obligations of the speaker’s language. They may be novel (at least for the speaker and/or the hearer), and to interpret them the hearer does not need to have had any prior experience of them. In a usage-based framework, no boundary is assumed between ‘rule-governed’ and ‘conventional’, nor need they be mutually exclusive. They can be thought of as being at either end of a cline of conventionality, running from novel sequences, through sequences recognized as regular and associated with a particular context and particular implicatures, to the most opaque idioms and ossifications which are unitary in that their meaning is impenetrable to those without prior experience of them. Conventionality in this sense depends on regularity, not rules. Given such a cline, it is not possible to delimit precisely the domain of phraseology. Moreover, what is conventional for one speaker might not be so conventional for another. In a panchronic usage-based framework, of course, rules themselves can be understood as having evolved over time through conventionalization.

Key to phraseology, then, is the notion that, for many linguistic sequences, there is more to the meaning of the sequence than the composition of the component parts. However, the nature of this “more” is uncertain. Overall, phraseology remains a rather ill-defined domain. The regular, “patterned” expressions it focuses on include collocations, idioms, and fixed phrases, with varying degrees and types of “additional meaning”. There is no obvious way to distinguish between idiomatic and non-idiomatic. Semi-lexically-filled regular chunks of language as well as unfilled schemata can also be viewed as phraseological. As Altenberg notes,

it is difficult to classify the many different types (Altenberg 1998: 101; cf. Ebeling & Ebeling 2013). From the structural point of view it is convenient to think of the various types as being along a cline from maximally filled, where both the words and the word order are fixed (e.g. idioms such as *when all's said and done* or *bubble and squeak*) through semi-filled/semi-schematic where some slots only are filled (e.g. 'have a (MODIF) time', as in *have a {terrible / splendid / absolutely riotous /..} time*), to maximally schematic where the lexical slots are unspecified (e.g. 'COMP and COMP', as in *faster and faster, higher and higher*).

As an umbrella term for phraseological types, 'phraseme' seems appropriate. Phrasemes crucially involve regularity, frequency, and conventionalization; through these mechanisms they come into being. Phraseology has tended to focus on the lexically-filled or semi-filled sequential structures or 'chunks' or 'lexical bundles' at sentence-constituent level. Above constituent level, towards sentence level and beyond, phrasemes tend to be more schematic, which goes along with decreasing predictability at "higher" levels of language production. Of course, schematic constructions with few or no lexical specifications can be observed both below and above sentence level. But at multi-clausal level, filled phrasemes become rarer and more specialized (this is the realm of proverbs, quotations and sayings). Larger discourse chunks can nonetheless be "phraseological" but in a more abstract sense, as exemplified below. Regular patterns, expressing discourse meanings such as discourse coherence or discourse salience, can be considered phraseological. At discourse level, as at other levels, different types of phraseme can be identified. They include maximally schematic structures (i.e. lexically unfilled), such as 'bareV-1 + NP-1 + bareV-2 + NP-2', as in *Build a dam, kill a river*, where syntactic form and clause order combine to convey the additional, phraseological meaning. Structures dependent on a connective, such as 'p CONN q' can also be considered phrasemes, and where the slot is filled, as in 'p so q', we have a semi-schematic phraseme. The 'consequence' meaning is dependent on the presence of the whole sequence. These observations are of course compatible with a Construction Grammar approach. In fact, observation of idiomaticity contributed greatly to the impetus for that approach (Michaelis 2019).

Our discussion of discourse-level phrasemes will use the relatively neutral term 'discourse pattern'. It will also use the term 'rhetorical pattern' for discussion of the rhetorical functions of such patterns. However, this is simply for convenience, and nothing hangs on the choice of terms. Discourse patterns serve to link sections of discourse together and to structure them into a salience hierarchy. In other words, they provide discourse coherence. As observed above, phrasemes can be situated on a cline of schematicity. Discourse patterns include those schemata where two (or more) ideas are linked by a third idea which is the coherence relation between them (the 'relational proposition' of Rhetorical Structure Theory (Mann

& Thompson 1986)). Coherence relations may be inferred or explicitly marked (by grammatical means such as morphosyntactic clause-combining devices or by lexical means such as connectives). In either case, regular patterns (discourse patterns) emerge and conventionalize over time.

2.2 The aims of the study

This study adopts the wide view of phraseology, to include recurrent discourse patterns, presented in 2.1, to compare the temporal expressions *then* and *now* and their frequent translation equivalents in French, *alors* and *maintenant*, in scripted and unscripted political discourse.

A question to be posed in the study is how ‘discourse patterns’ compare across languages; to what extent different languages conventionalize similar patterns using comparable grammatical and lexical means for comparable discourse meanings. The starting point is the recruitment of temporal expressions for discourse-structuring functions: the comparison involves English *now* and *then* and French *maintenant* and *alors*. The study aims to compare the polysemies and the usages of these pairs of time expressions, which can be considered translation equivalents in both their temporal and their extended senses, to see how far the parallels between the two languages can be taken. The comparison is made from a discourse-level point of view, taking into account the discourse-structuring role of the adverbs. When comparing, or translating between, related languages, such as English and French, the unit of comparison adopted is often the sentence, defined as a syntactic unit. The syntactic sentence continues to enjoy a special status in linguistic analysis as the default unit of discourse. But such an approach risks undervaluing the role of a great deal of structure beyond the sentence, both referential (e.g. anaphoric/cataphoric) and rhetorical (including information structure and coherence relations).

A second question to be considered is the role of genre: to what extent are genre differences similar across languages? This consideration may suggest hypotheses regarding which aspects of discourse structuring are potentially universal and which depend on language or local convention.

3. Data

3.1 Comparable corpora

Much contrastive analysis is based on bidirectional parallel corpora, i.e. translated texts. Translated text has the advantage, among others, of allowing equivalences across the different languages to be quickly identified, insofar as the ‘same’ meanings are expressed in the two languages. The meaning of the discourse acts as a *tertium*

comparationis. But basing contrastive analysis on translated text also has disadvantages, such as its dependence on translation type and translation quality. Contrastive analysis at the suprasentential level is not so easy with parallel corpora, partly because at discourse level the distinctions between languages are less grammatical and more rhetorical than at lexical or clausal levels; translation choice is therefore less constrained. Atypical rhetorical features are often what gives translated text its foreign flavour. In translation it is not easy to recreate in the target language the links between particular genres and particular discourse structures. Comparable corpora, containing genre-matched native text, have to take the comparable situations as the *tertium comparationis*, and situations are never entirely comparable. The advantage is that the calquing of discourse features and structures is avoided, meaning that the identification of a discourse feature as characteristic of a genre is more reliable because the frequencies and distributions are native.

3.2 The political discourse corpora

This study is based on two small comparable corpora of political discourse, one of scripted monologic political speeches, one of unscripted dialogic political interviews and discussion. The assumption is that the event types of political speech and radio or TV political discussion programme in France and the UK are comparable. The corpora are described in Table 1.

Table 1. The comparable corpora used in this study

Political speeches genre (scripted monologue)	Political discussion genre (unscripted dialogue)
French part: 383,800 words	French part: 159,000 words
148 political speeches, 1–4 per speaker 53 speakers 2001–2010	57 discussions and interviews 45 speakers (politicians and political commentators with journalists) 2000–2017
English part: 385,700 words	English part: 140,100 words
132 political speeches, 1–5 per speaker 32 speakers 1995–2010	59 discussions and interviews 71 speakers (politicians and political commentators with journalists) 2004–2009

The political speeches corpus consists of speeches given by politicians, mostly government ministers. In the political contexts of France and the UK, the genre of ministerial speeches is relatively constrained and relatively comparable. The situations in which such speeches are produced are well-defined and similar across the two countries, and identifying comparable texts for a corpus is fairly

straightforward. The speeches are pre-scripted and made available to journalists and to the wider public. The speeches are designed for a wide audience of other politicians, other governments, other institutions, the media and the public, in addition to the immediate audience that is physically present. Speeches are written to be spoken and comprise a few thousand words at most. Speeches have several functions: expository, persuasive and ceremonial, in varying measures depending on the context of production of the speech and on the conventions and political traditions of the country. Speeches are integral parts of many regular events in the calendars of politicians.

The political interviews and discussion corpus is a little less constrained. The dialogues come from radio and television programmes in which politicians and other personalities involved in politics are interviewed in one-to-one situations, or participate, under the direction of a programme presenter, in discussion with a small group of people. Some of the politicians that are present in the speeches corpus also appear in the discussion programmes. The corpus is based on transcripts made available by the media that broadcast the programmes. They show no prosodic features and no doubt edit out hesitation markers and other non-linguistic material but appear to be relatively reliable records of the words spoken. The examples from the interviews and discussion corpus are presented here with the punctuation removed and replaced by dots to avoid a false impression of written text. Square brackets ([...]) signal that a long passage has been abbreviated.

'PS' refers, in the source information following the examples, to the political speeches comparable corpus, while 'PID' refers to the political interviews and discussion comparable corpus.

4. *Now* and *maintenant*

4.1 Polysemy and discourse patterns

As mentioned in the Introduction (Section 1.1), time expressions often develop argumentational or text-organizational functions. This is the case for *now* and *maintenant*, which both have temporal and argumentational uses.

The examples in (1) illustrate temporal *now* and *maintenant* in medial position (i.e. post-V for French and post-Aux pre-V for English), the most frequent position for both languages.

- (1) a. *some of the country's best known businesses are now agreeing to take part in the new deal project* (PS, 1997)
- b. *je crois que la Corse a maintenant besoin d'action beaucoup plus que de discours ...* (PS, 2002)
 ['I believe Corsica **now** needs action much more than speeches']

Argumentational *now* and *maintenant*, in initial position in English and French, serve the speaker's rhetorical-structuring purposes as topic shifters. They signal to the listener or reader that what is coming up is a new idea that departs from what has gone before in being a new subtopic, a shift to a new aspect or development of the overall topic, or a different perspective on the topic, setting it in a new and different light. There is thus always a contrast present in the context, reminiscent of the usage of the initial temporal adverbs, but this time a rhetorical contrast. Example (2) illustrates how English and French use *now* and *maintenant* to introduce a different perspective, in these cases the speaker's perspective, on the current topic. In both languages the adverb is in initial position for this function and serves a presentational function, putting into focus the idea it precedes.

- (2) a. *you can look at any one of these things.... now ... what I have tried to do is I've tried at all times to do what the rules required* (PID, 2009)
- b. *je parle d'incertitude et de demande de clarification [...five turns...] maintenant... quand je parle d'incertitude je fais simplement un constat et ce n'est pas un jugement de valeurs* (PID, 2016)
- [I'm talking about uncertainty and about requests for clarification [...five turns...] **now**... when I talk about uncertainty I'm simply making an observation and not a value judgment ...]

As with many similar polysemous expressions, where a more abstract use has emerged from a more concrete one, there is evidence of persistence in both *now* and *maintenant*. It is thus usually possible to understand argumentational uses such as those in (2) both as referring to the time of speaking ('At present I say...') and as marking a contrast ('A new point is that ...').

In the corpora, *now* and *maintenant* show similar patterns to one another in their temporal usage. Both enter into a contrastive discourse pattern in initial position (3):

- (3) a. *when in the old days you would have gone to the post office to get ... renew your car tax ... you'd have had to sort out your MOT certificate ... your log book... your insurance... all of those things... queue up in the post office with a cheque now you can do it online or by phone.* (PID, 2009)
- b. *avant ... les quatre grands pays ... l'Allemagne ... la France... l'Angleterre... l'Italie... avaient 10 voix chacun ... maintenant ... avec la repondération... elles ont 29 voix* (PID, 2000)
- [**before** ... the four largest countries ... Germany... France... the UK... Italy... each had 10 votes... **now** they have 29 votes.]

In both (3a) and (3b) there is a contrast between a time in the past and the present. Such juxtaposition of two contrasting ideas, framed by past and present temporal expressions, forms a schematic rhetorical pattern '[past time expression/tense] p'

‘[present-time expression] *q*’. The result is that temporal *now* or *maintenant* in initial position tends always to evoke a contrast.

The corpora also reveal a more complex pattern with *now*, where, alone or in combination with another marker, it participates in an argumentational sequence. The examples in (4) illustrate this for concession, where a claim is followed by a concession (which may be elaborated by an explanation or justification or evidence as in (4a)) and then a claim which constitutes the speaker’s main point, and which is consonant with the first claim. The pattern can be schematically represented as ‘[claim] [concession marker(s)] [concession (elaboration)] [contrast marker(s)] [main claim / counterclaim]’. The concession is made explicit in (4b, c) by *of course*, while *now* marks the introduction of a different perspective, one that the speaker does not wish to emphasize.

- (4) a. *what we did was to create a governing body which included Central Office ... the parliamentary party and the voluntary party and they have worked very very well together ... now nobody has pretended that those rules were absolutely unchangeable ... in fact we built into the Constitution a provision to review things after a period.... but the truth of the matter is... they did succeed in uniting the party.* (PID, 2005)
- b. *the government [...] had the chance to stop it... but that was back in October... now of course they should look at all the legal avenues that exist to try and make sure that not all of this seven hundred thousand pounds is paid to Fred Goodwin...but this is a bit like trying to bolt the stable door after the horse has itself bolted.* (PID, 2009)
- c. *I think that I think that by allying ourselves with the United States [...] we sort of threw our hat into the wrong ring I believe...now of course I don’t excuse what these people did...but I think the way it works is that [...]* (PID, 2005)

The markers (*now* and *but* in (4)) together form a structure or frame for the concession and its counterclaim. Even where there is no obvious concession in the content, the schematic structure suggests a concessive argument: in this structure, *now* marks its host as given or backgrounded in some way or as a premise for a claim, and sets up an expectation of a rhetorically strong claim to follow.

Maintenant was not found to participate in any wider regular pattern or in any regular collocations, perhaps because of its relatively low frequency as an argumentational marker in the data, or because it has developed an argumentational function relatively recently. The more grammaticalized *or* (from temporal *ore/ores*, ‘now’, from Latin *hora*) does show a pattern in the data, but unlike *now*, it marks

its host as foregrounded and in opposition or antithetical to the previous idea, as exemplified by the two different speakers in (5).

- (5) a. *ce que je regrette un peu c'est que le Président de la République [...] n'ait pas anticipé les conséquences ... or nous étions un certain nombre à les voir* (PID, 2000)
 ['what I rather regret is that the President of the Republic [...]
 didn't anticipate the consequences ... or several of us saw them']
- b. *là où à mon avis les choses ne vont pas c'est dans le fait que le Président de la République n'accepte pas l'idée d'une crise de la politique ... or il y a une crise de la politique* (PID, 2000)
 ['where in my view things are not right is that the President of the Republic doesn't accept the idea of a political crisis ... or there is a political crisis']

So while *now* and *maintenant* (and *or*) have similar polysemy structures insofar as they are temporal and argumentational, the argumentational functions have taken a different direction for each expression.

4.2 Frequency comparison

By contrast with the similarities between the polysemies, the frequencies reveal considerable differences between the French and English expressions. Figures 1 and 2 compare the frequencies for the scripted, monologic political speeches genre and the unscripted, dialogic political interviews and discussion genre respectively.

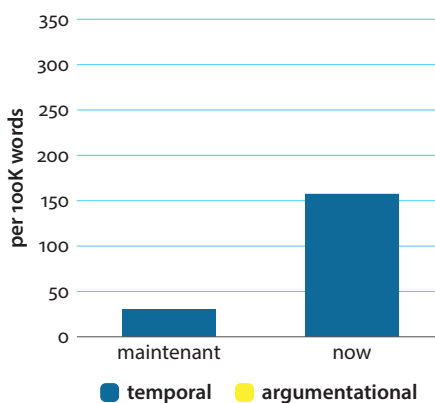


Figure 1. *Maintenant* and *now* in the political speeches corpus

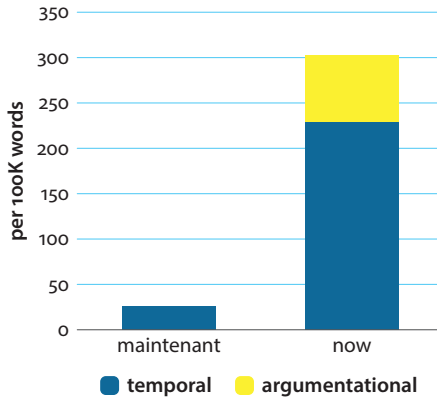


Figure 2. *Maintenant* and *now* in the political discussion corpus

First, for both genres and both languages, the temporal use dominates, consistent with the argumentational uses having developed relatively recently. For the dialogue corpus, the proportions are similar across English and French.

Second, there is an important genre difference: for both languages, the argumentational use occurs predominantly in the more informal, unscripted, dialogic data. For English, the frequency of argumentational *now* was 1 per 100K words in the speeches, but 71 per 100K words in the spontaneous discourse of the interviews and discussions. The French speeches contained no examples at all of argumentational *maintenant*. This could be partially due to the availability of the more formal *or*, which grammaticalized earlier (sixteenth century). In the speeches, however, *or* frequency is lower, at 8 per 100 K words, than in the dialogic corpus (15 per 100K words), and, as seen in Section 4.1, its function is rather different.

Turning to the relative frequencies, in the dialogic corpus, argumentational *maintenant* represents just 5 per cent of total occurrences, whereas argumentational *now* accounts for 31 per cent of occurrences. It is possible that the relative frequency of argumentational *now* has been boosted by its participation in regular schematic rhetorical patterns as outlined in 4.1.

5. *Then* and *alors*

5.1 Polysemy and discourse patterns

Like *now* and *maintenant*, *then* and *alors* have similar polysemies. Both can signal ‘at that time’ with reference to some past time already established in the discourse (6). But whereas *then* is widely used in the sense of ‘next’ or ‘after that’ for both past (7) and future situations, *alors* in the data rarely is.

- (6) a. *In 1979 there were 59,000 full-time higher-education students in this country, many of them simply enjoying the very low fees we then charged.* (PS, 1995)
- b. *elle intégrait l'information dont nous disposions alors sur le ralentissement économique constaté au 1er semestre* (PS, 2001)
 ['it included the information we **then** had on the economic slowdown in the first semester']
- (7) *We had the oil crisis... then we had the credit crunch* (PID, 2008)

And like *now* and *maintenant*, *then* and *alors* have also developed argumentational uses. And again, there is evidence of persistence of the temporal sense in the newer uses.

Both *then* and *alors* contribute to two related rhetorical patterns, which we will term 'inferential' (following Quirk et al. 1985) and 'conditional'. The argumentational function, in both cases, presupposes the temporal function insofar as the argumentation links two situations that are necessarily temporally sequential, as cause precedes consequence. From the diachronic perspective, one event occurring after another related event becomes associated with causation, whence the cross-linguistically common temporal > causal semantic shift (Heine & Kuteva 2002).

Both *then* and *alors* function as inferential connectives referring anaphorically to a previous idea as having resulted in the upcoming idea (8). In both (8a) and (8b) the expressions can be glossed as 'given that that is so' or 'since that is the case'; thus the idea in the first unit in each example is presented as providing the justification for the proposal in the second. While other common markers such as English *so* and French *donc* mark a wide range of consequence, *then* and *alors* appear in the data where the speaker is justifying a conclusion by reasoning, by inferencing.

- (8) a. *you're quite right ... that we have our own problem with one dodgy donor who gave money to the party. Let's then take the opportunity to change this* (PID, 2009)
- b. *il y a des points de vue philosophiques et des points de vue religieux qui sont différents... ils sont respectables... alors acceptons le débat* (PID, 2012)
 ['there are philosophical points of view and religious points of view that are different... they can be respected... **then** let's agree to discuss the issue']

The second pattern is centred around conditional constructions. The redundant inclusion of *then* in the English conditional pattern 'if p, (then) q' and of *alors* in

the French one '*si p (alors) q*' may be "mannered", but is common in the political discourse of both corpora, illustrated in (9).

- (9) a. *If we do that, then Africa has a chance to ...* (PID, 2008)
 b. *Si nous ne parvenons pas à créer cette citoyenneté européenne alors c'est la civilisation européenne qui sera bousculée par la mondialisation.* (PID, 2011)
 ['If we do not manage to create this European citizenry then European civilization will be overturned by globalization.']

Although they are redundant for the expression of the condition, English *then* and French *alors* arguably can serve other purposes. First, they can contribute to emphatic information structure, acting as presentationals with the rhetorical function of focusing the idea that follows (Hansen 1996: 141–142 on the foregrounding function of *alors*). They may also have a further rhetorical function of creating a rhythm. The role of rhythm in public speaking is recognized in traditional rhetoric, as explained for example in Smart's 1848 manual (where protasis and apodosis refer not only to conditional constructions but to subordinate constructions more generally):

If [...] the subject and occasion require a sustained style; a style in which every 'protasis' raises expectation, and every apodosis fulfils it; we shall be wanting in powers of language, should the ear suggest nothing higher in rhythm and construction, than we find in sentences of ordinary occurrence. (Smart 1848: 33)

Prosodic and formal parallelism of the *if-then* type provides such rhythms. In the data, *then* and *alors* also enter into parallelisms with other temporal expressions such as *when*, *as soon as*, *quand*, *lorsque* (10), where there is equally a strong implicature that the second idea is presented as being a consequence of the first:

- (10) a. *When you see the financial industry caving in and doing things that Nick Leeson got six and a half months for... then I think Brown's got the message.* (PID, 2008)
 b. *Lorsque nous aurons fait cette union... alors... effectivement... il y aura possibilité de mutualiser la dette.* (PID, 2011)
 ['When we have created this union... then... in fact... it will be possible to mutualize the debt.']

The structure is also found in conjunction with other markers including *unless*, *as*, *where*, *because* (11). As with the *if*-conditionals, the *then* or the *alors* is semantically redundant. But it serves an information structuring function by marking the start of, and putting into focus, the speaker's main, concluding point. The first idea

is subordinated or backgrounded by the first marker and the second is focused by *then* / *alors*. In the examples of (9), (10) and (11) the markers also provide rhythm to the whole, especially clearly in the extended *because-then* structure in (11b).

- (11) a. *Where Tony Blair has picked up Conservative ideas, then we're not going to abolish them ...* (PID, 2006)
- b. *Because he's got a sense of humour...because he's human...because he's likeable... because he doesn't fit the standard political mould...then I think that people within the political establishment say ah you know he's riding for a fall.* (PID, 2008)

These constructions can be seen as rhetorical elaborations of the inferential use. Some occurrences of *then*, however, seem not quite to fit the inferential or conditional semantic mould. This is the case for (10a), where there seems to be some incoherence.¹ That is, it is hard to see how the second idea could be a consequence or an inference of the first. Nor does it make better sense interpreted as temporal. A possible explanation is that the speaker can take advantage of the regular inferential use of the construction to suggest a line of valid reasoning where there is none. At the same time, it is possible that the combination of the regularity and the rhythm of rhetorical patterns acts as a kind of mnemonic for the speaker. That is, the speaker first reaches for a pattern, such as 'justification + claim', framed by associated backgrounding and foregrounding information-structuring markers, and then fills in with content. Such patterns recall what Pawley calls "productive speech formulas": "clause- or multi-clause-sized constructions that contain some slots that are lexically specified and others that are filled by abstract grammatical categories" (Pawley 2009: 19), but rather than abstract grammatical categories, they are filled by rhetorical types such as 'evidence', 'concession' and so on.

The importance of rhythm to "professional" speakers of specialized genres is emphasized by Kuiper (2000), who suggests that speakers use formulas and their rhythms to mitigate the effect of short-term memory constraints on fluency and complexity. The examples in (11) also come across as odd, due to the redundancy and unconventionality in the '*because p then q*' construction, but they make sense when seen as a speaker's mnemonics, where the rhetorical value of the markers overrides their semantic value.

1. Cases such as (10) or (11) are not (yet) directly comparable with so-called 'Austrian conditionals', which are conventionalized uses of the English *if*-construction (one type of insubordination).

The case of French *alors* is rather different. And *alors* appears to have gone a lot further along the path towards presentational functions than *then*, often occurring as nothing more than a focus particle. For some speakers, it regularly appears at the start of an answer to a question (12) (here it is comparable to English *well*). Inter-speaker variation in frequency suggests a very bleached marker. For some speakers a very bleached marker can become a feature of their style or even a kind of “linguistic tic”. For example, the speaker in (12) starts almost half their answers to the interviewers’ questions with *alors*; other speakers rarely do.

- (12) a. Q -*On évoque maintenant un éventuel retour en arrière... où en est-on?*
 A -*Alors on regrette cette décision britannique...* (PID, 2017)
 [‘Q -There’s talk now of a possible backtracking... what’s the situation?’]
 [‘A -*Alors*... we regret this decision by the British’]
- b. Q -*Est-ce que vous avez des nouvelles de Loup Bureau?*
 A -*Alors on suit sa situation très attentivement* (PID, 2017)
 [‘Q -Do you have any news of Loup Bureau?’]
 [‘A -*Alors*... we are monitoring his situation very carefully’]

The English data *well* functions in a very similar fashion. Example (13) is taken from one discussion between an interviewer and three guest interviewees and shows one response from each interviewee; all three repeatedly begin their answers with *well*.

- (13) Q: *your thoughts on the events of the last few days first of all*
 A1: *well I just heard from America ...*
 Q: *what are you thinking about doing?*
 A2: *well first of all I think it’s important to point out that this is a free vote issue ...*
 Q: *why don’t we encourage that?*
 A3: *well I think the government will have to encourage more saving ...*
 (PID, 2004)

Alors occasionally occurs discourse-initially, as in (14a) which is the first turn of the presenter at the start of a programme, to introduce the topic. This is not a function of *then*. But there is an interesting parallel with English *now*, in its topic-changing, topic-introducing function: (14b) is from the start of a political discussion programme where the presenter introduces the first topic.

- (14) a. *Bonjour... alors... l'un des quatre policiers attaqués samedi...*
(PID, 2016)
[‘Good morning... *alors* one of the four policemen attacked on Saturday...’]
- b. *We have a live studio audience who will be debating with a stellar political panel [NAMES]... now Labour are way behind in the opinion polls...*
(PID, 2006)

Other discourse-initial topic introducers found in the English data are *well* and *so*.

There are 19 examples in the corpus of a more bleached *then* used as an additive to introduce an additional or further point (15). Most are preceded by *and* or *but*, with which *then* combines to form a complex additive/conjunction.

- (15) *[we] make sure that the facilities and the coaching support is there for them and then of course I mean we've done an enormous amount of development for facilities for elite athletes*
(PID 2004)

Hansen (1995) discusses how *puis* (unlike *alors*) has undergone grammaticalization into an additive marker and possibly into a conjunction (1995:53). This is borne out by our French data, where there is evidence that *et puis* has grammaticalized to additive/conjunction (16).

- (16) *Vous avez dit que vous aviez soutenu et puis maintenant vous ne soutenez plus...*
(PID, 2016)
[‘You said that you did support it *et puis* now you no longer support it ...’]

The next sub-section looks at differences in frequency, across the genres and the two languages.

5.2 Frequency comparison

Figure 3 compares the frequencies in the monologic speeches of the temporal, conditional, inferential, presentational and additive functions of *alors* and *then*, and Figure 4 compares the same functions for the dialogic corpus.

What is similar across the two languages is that the frequency of *then* and *alors* is much greater in the dialogic language than in the scripted monologic speeches: three times greater in the case of *then*, and six times greater for *alors*.

But the distributions are very different. While the proportions of the different uses of *then* are rather similar across the two genres, this is not at all the case for *alors*, where temporal use is lower in the dialogic data, the vastly greater frequency being accounted for almost entirely by inferential and presentational uses. A similar but less striking observation was made for *now* and *maintenant* (4.2).

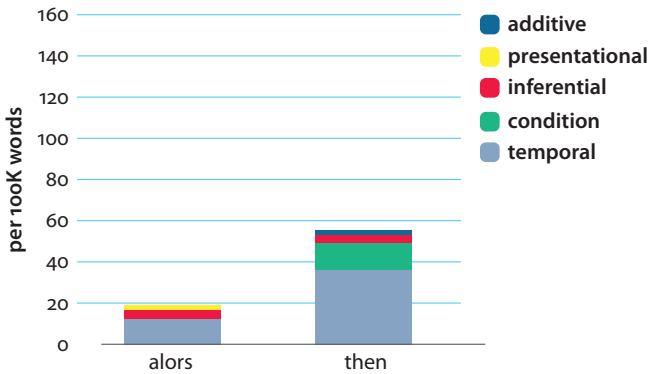


Figure 3. *Alors* and *then* in the political speeches corpus

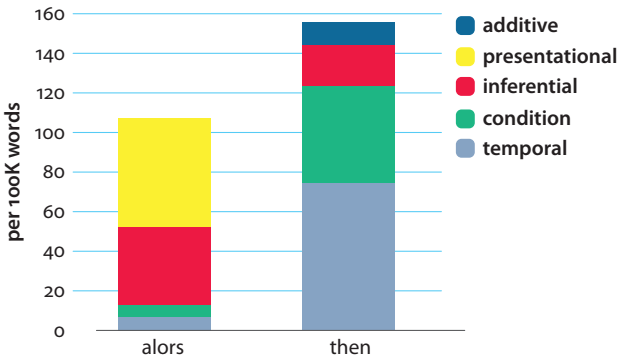


Figure 4. *Alors* and *then* in the political discussion corpus

6. Conclusion

On the face of it, the polysemy structures of the pairs of English and French adverbs as seen in these corpora are rather similar. The adverbs exemplify the cross-linguistic tendency for temporal adverbs to grammaticalize into markers of rhetorical relations. Each is used in both temporal and argumentational senses. Moreover, in both English and French the expressions occur in initial, medial and final positions. All four adverbs typically occur in the standard adverbial positions for each language when temporal, and utterance-initially in both languages in their rhetorical function. *Alors* and *then*, however, have been shown to be developing final-position uses in their rhetorical function (see Degand & Fagard 2011 for *alors*; Haselow 2011 for *then*). *Now* and *maintenant* have both evolved discourse-structuring uses; they occur in contrastive contexts and serve topic-management

functions. *Then* and *alors* have likewise developed into discourse markers of inference, alone and in conditional contexts; both have become markers of information structure, foregrounding or focusing the upcoming idea relative to a previous idea.

There are nevertheless considerable differences between the English adverbs and the French ones. First, they seem to be at different stages of grammaticalization. *Maintenant* in the corpus data is less frequent and less grammaticalized than *now*, (possibly due to earlier-grammaticalized *or* still occupying this space in formal genres). While *now* is seen to recur in a concessive discourse pattern, collocating with other markers, no such pattern was apparent for *maintenant*. In the case of *alors* and *then*, the French lexeme occurs in a wider range of rhetorical contexts, and looks more bleached and grammaticalized than *then*, despite its being less frequent in these genres (cf. Degand & Fagard (2011) and Haselow (2011) on functional splitting in *alors* and *then* respectively). The two adverbs function in similar ways in conditional and inferential discourse contexts, but differ in discourse management, *alors* having wide presentational functions while *then* can introduce an additional idea.

Second, the argumentational uses of the English and French adverbs develop along similar lines, such as the inferential senses of both *then* and *alors*, which retain some of the temporal sense. But once the adverbs acquire more abstract discourse functions, they no longer occur in such comparable contexts, and they cannot be taken for translation equivalents. As has been seen, *then* occurs as an additive while *alors* can “present” almost any statement or question.

Striking differences appear when frequency and genre are taken into account. For *then/alors*, considerable genre difference is apparent for both languages, with much greater frequency in the dialogic data. But the difference is much greater in the French data compared with the English. This is particularly interesting given that several of the speakers are represented in both corpora. It suggests that in the two genres of political speeches and political discussion, usage of the two French adverbs is quite tightly bound to particular discourse patterns that are semi-conventionalized. And it may reflect a wider register gap in French than in English between the standard formal language of the written-to-be-read speeches and the spoken language of unscripted discussion.²

2. In France, “alongside the national standard language, is a nationwide, non-regional non-standard *français populaire* ... now part of virtually every Frenchman’s linguistic repertoire” (Posner 1997: 74, emphasis added). Posner describes it as a case of “stable variation between standard and non-standard, possibly throughout the modern period” (1997: 75). Educated French speakers such as those represented in the corpus command both the standard (written) variety and the nationwide non-standard variety which is the norm for speech. This is a simplification of the complex register situation of European French, of course, but it highlights the

The development of sub-sentential phrasemes involves the amalgamation or coalescence of meaning into the sequence, resulting in a loss of compositionality. There is likewise coalescence at discourse level: coherence-relational “meaning” cannot be said to reside only in the coherence markers, but in the whole pattern or schema of which it forms part. There are some signs that such patterns may have a mnemonic value for the speakers, whereby those accustomed to public speaking develop a repertoire of semi-automated rhetorical patterns, like templates, with their associated prosodies and information structures, into which they slot the relevant claims and arguments. The aim of much political language as public discourse is to convince by (apparently) reasoned argumentation, involving causation, concession, condition, justification, evidence, example and so on. Markers such as these temporal adverbs help form a linguistic framework for such argumentation. *Now*, *then* and *alors* appear to participate in some of these rhetorical patterns, while *maintenant* does not. The data also suggested that while the argumentational uses in English and French follow broadly parallel ways, motivated by aspects of the temporal meanings, further bleaching of the expressions into information-structure functions such as focus was more arbitrary.

An important factor in the ongoing development and polysemy of the four adverbs that has not been addressed here is their role in complex subordinators (such as *now that* or *alors que*, etc.) and the way they combine with other markers of coherence and/or information structure. A further issue that needs to be addressed is the absence of any comparison with other, more frequent genres. The dialogic data need to be put in the context of domain-general conversational language, and the speeches compared with other written and written-to-be-spoken genres. Only then can the interaction between frequencies and polysemies be appreciated. The study has shown nevertheless that there are strong genre effects on the usage of the adverbs and that these effects operate in similar directions across the two languages.

Acknowledgments

Many thanks to the editors of this volume, to the participants and audience at the workshop “Languages in time, time in languages: Phraseological perspectives” at ICAME-40 in Neuchâtel in 2019 and to two anonymous reviewers.

contrast with the situation in the United Kingdom, where there is no nationwide, non-regional non-standard *anglais populaire* (despite some dialect levelling, non-standard British English remains regional (Beal 2010)) and where morphosyntactic and stylistic differences between the “standard” or “educated” spoken language and the written standard are much less marked.

References

- Altenberg, Bengt. 1998. On the phraseology of spoken English. The evidence of recurrent word combinations. In *Phraseology, Theory, Analysis and Applications*, Anthony P. Cowie (ed.), 101–122. Oxford: OUP.
- Degand, Liesbeth & Fagard, Benjamin. 2011. *Alors* between discourse and grammar. The role of syntactic position. *Functions of Language* 18: 29–56. <https://doi.org/10.1075/fol.18.1.02deg>
- Ebeling, Jarle & Ebeling, Signe Oksefjell. 2013. *Patterns in Contrast* [Studies in Corpus Linguistics 58]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.58>
- Hansen, Maj-Britt Mosegaard. 1995. *Puis* in spoken French: From time adjunct to additive conjunct? *Journal of French Language Studies* 5(1): 31–56. <https://doi.org/10.1017/S0959269500002490>
- Hansen, Maj-Britt Mosegaard. 1996. Some common discourse particles in spoken French. In *Le Discours: Cohérence et Connexion* [Études Romanes 35], Maj-Britt Mosegaard-Hansen & Gunver Skytte (eds), 105–149. Copenhagen: Museum Tusculanum Forlag.
- Haselow, Alexander. 2011. Discourse marker and modal particle: The functions of utterance-final *then* in spoken English. *Journal of Pragmatics* 43(14): 3603–3623. <https://doi.org/10.1016/j.pragma.2011.09.002>
- Heine, Bernd & Kuteva, Tania. 2002. *World Lexicon of Grammaticalization*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511613463>
- Kuiper, Koenraad. 2000. On the linguistic properties of formulaic speech. *Oral Tradition* 15(2): 279–305.
- Mann, William C. & Thompson, Sandra A. 1986. Relational propositions in discourse. *Discourse Processes* 9(1): 57–90. <https://doi.org/10.1080/01638538609544632>
- Michaelis, Laura A. 2019. Constructions are patterns and so are fixed expressions. In *Patterns in Language and Linguistics*, Beatrix Busse & Ruth Moehlig-Falke (eds), 193–220. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110596656-008>
- Pawley, Andrew. 2009. Grammarians' languages vs. humanists' languages and the place of speech act formulas in models of linguistic competence. In *Formulaic Language, Vol. 1: Distribution and Historical Change* [Typological Studies in Language 82], Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen Wheatley (eds), 3–26. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.82.01gra>
- Posner, Rebecca. 1997. *Linguistic Change in French*. Oxford: The Clarendon Press.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Smart, Benjamin H. 1848. *A Manual of Rhetoric*. London: Longman, Brown, Green, and Longmans.

Time adverbials in English and Norwegian news discourse

Hilde Hasselgård

University of Oslo

This chapter examines time adverbials in English and Norwegian as evidenced in two corpora of news articles. The adverbials are identified through a manual, bottom-up procedure, and their syntactic realizations, semantic types and positions are analysed. The comparison includes the lexical realizations of time adverbials in both languages, partly through the lens of lexical priming. Similarities between the languages include the distribution of syntactic and semantic types of time adverbials. Cross-linguistic differences in adverbial placement are evident in clause-medial position, where English is more restrictive than Norwegian. The lexical comparison shows that the languages may differ in how similar meanings are realized. Furthermore, the lexical priming of some frequent lexical items reveals lexeme-specific and possibly register-specific patterns.

1. Introduction

Time adverbials are a pervasive feature of news discourse, which can be related to the communicative “focus on reporting current events” (Biber & Conrad 2009: 119). As in narrative text types, the use of time adverbials in news texts can serve to narrate sequences of events as well as (inferable) cause-effect relationships (Biber & Conrad 2009; Virtanen 2014). A special feature of newspaper texts is that they are normally designed to be read on the day of publication, which enables meaningful use and interpretation of deictic adverbials such as *yesterday*, *today*, and *last week*. Hoey (2005: 132) goes as far as to claim that “[n]ewspapers are more aware of their place in time than any other kind of discourse.” Example (1), in which the time adverbials have been highlighted, may serve as an initial illustration:

- (1) RETIRED factory worker Ted Whan's life has been made hell by queues of bendy buses blocking his drive. The 75-year-old of Stornoway Road, Castle Vale, says he has suffered abuse and intimidation from drivers, had his car hit *twice* and *often* cannot get in or out of his garage. He had enjoyed free access to his garage, off Turnhouse Avenue, *for 35 years*, but *since last year* has *more often than not* had to wait for a bus to move *before he can come and go*. (BE06: A20)

This chapter presents a contrastive study of time adverbials in English and Norwegian news reports, based on comparable corpora (see Section 3). One aim is to survey cross-linguistic similarities and differences in syntactic and semantic types of time adverbials as well as in their placement. The framework for classifying and analysing adverbials and their positions relies on Hasselgård (2010); see also Biber et al. (1999). A second aim is to investigate recurrent lexical realizations of time adverbials with regard to their phraseological make-up and their potential patterns of lexical priming (Hoey 2005). To this end, I will investigate the collocational and colligational (syntactic) behavior of selected recurrent adverbials to see whether (and how) the patterns of similar lexicogrammatical items vary across the languages.

The research questions are as follows: What are the differences and similarities between time adverbials in English and Norwegian news discourse in terms of semantic types, syntactic realizations, and positions? What are the most common lexical realizations of time adverbials in both languages? Are there cross-linguistic differences in the lexical priming of frequent time adverbials in the news report register?

Temporal expressions have been found to be more frequent in Norwegian than in English (Ebeling & Ebeling 2017; Hasselgård 2017a). Although these previous studies have mainly been based on fictional texts, the expectation is that time adverbials will be more frequent in the Norwegian news reports, too. Adverbial placement is known to differ between Norwegian and English, and also to be register-dependent. (Biber et al. 1999; Hasselgård 2010, 2014a). It is therefore expected that initial and medial placement will be more frequent in Norwegian than in English (see further Section 2). The frequent use of deictic time adverbials in English newspapers was noted in Hasselgård (2010: 105), so this feature is expected to come up in the present material as well. However, it is difficult to formulate further hypotheses on other aspects of the lexical realizations of adverbials, as there are no previous studies of this.

In the remainder of this chapter, Section 2 outlines some relevant previous research, and Section 3 describes the material and the method used in the study. Section 4 describes the classificatory framework for the analysis before the corpus findings are presented in Section 5. Section 6 offers some further discussion and concluding remarks.

2. Some previous work on time adverbials

Much has been written about the linguistic expression of time in general and temporal adverbials in particular. According to the framework of Quirk et al. (1985), time adverbials belong to the category of ‘adjuncts’ (as opposed to ‘disjuncts’, ‘subjuncts’ and ‘conjuncts’; see Quirk et al. 1985: 504). In Biber et al. (1999) this category corresponds to ‘circumstance adverbials’ (as opposed to stance and linking adverbials). Circumstance adverbials are more integrated into the clause structure than the other two types, and “add information about the action or the state described in the clause, answering questions such as ‘How, When, Where, How much, To what extent?’ and ‘Why?’” (Biber et al. 1999: 763). Various studies have found time adverbials to be among the most frequent adverbial types, e.g. Biber et al. (1999: 983 ff.) and Hasselgård (2010: 34).

Time adverbials are a highly heterogeneous category in terms of syntactic realization and positional flexibility; see Section 4. They furthermore serve a variety of functions at clause level by locating events and situations in time. At discourse level, time adverbials may function as text-structuring devices, marking coherence and indicating text segmentation (Virtanen 2014: 106). This function is most prominent in clause-initial position (*ibid.*: 107), and is moreover “particularly visible in narrative passages where events may be structured along a time axis” (Hasselgård 2010: 216).

The above observations all come from studies of English, but it has been shown that clause-initial adverbials can be discourse-structuring devices in Norwegian too (e.g. Hasselgård 2014a). In the news material examined in Hasselgård (2014a), time adverbials have this function more often than other types of adjunct. In fact, Norwegian was found to make use of this resource more often than English (*ibid.*: 89), which correlated with a higher frequency of time relationship adjuncts (*ibid.*) as well as other time adverbials containing anaphoric elements (*ibid.*: 88). Hasselgård (2014a) studied primarily clause-initial adverbials; however, it was noted that the overall frequencies of adverbials, in all positions, were higher in Norwegian than in English in both fiction and news (*ibid.*: 79).

Other contrastive studies that discuss adverbials and/or other temporal expressions in English and Norwegian include Ebeling and Ebeling (2017) and Hasselgård (2017a, 2017b). Ebeling and Ebeling, using a parallel corpus of fictional texts, find that recurrent n-grams with temporal meaning are significantly more frequent in Norwegian than in English (2017: 22). Similarly, Hasselgård (2017a) shows that a larger proportion of recurrent n-grams are temporal in Norwegian than in English. It is observed that “a temporal n-gram is more likely to be clausal in Norwegian than in English, which might point in the direction of higher frequencies of both temporal adverbial clauses and time adverbs,” since clausal n-grams are often identified as temporal by the presence of a time adverb (2017a: 96 f.).

Hasselgård (2017b) studies the placement of adverbial clauses, including temporal ones, in English and Norwegian fiction and news. The two languages are found to be rather similar in this respect, preferring end position for most clause types (conditional ones being an exception), with initial position as the second-most frequent choice and very little use of medial position. The choice between initial and end position in both languages is linked to information structure: “adverbial clauses containing anchored information are more likely to be sentence-initial, and those with new information are more likely to be sentence-final” (2017b: 137).

However, syntactic realization and semantic category are also important: non-finite clauses are less likely than finite ones to be initial, and only clauses denoting time and contingency were sufficiently frequent in both initial and final position to be subjected to a study of positional alternation. A problem of the 2017b study was that the material was very small. Hence, the present study uses slightly larger corpora, but see Section 3 for limitations.

To my knowledge, there are no usage-based, functional studies of time adverbials in Norwegian. However, the description of adverbials in Norwegian grammars such as Faarlund et al. (1997) and Holmes and Enger (2018) make it plausible that semantic and syntactic features of adverbials can be described in the same way for both languages in spite of the different classificatory frameworks in the reference grammars available. Further support for this is offered by Lindquist (1989), who applies the classificatory system of Quirk et al. (1985) to Swedish adverbials with only minor adaptations.¹ Note, however, that adverbial placement in Norwegian tends to be described in terms of a positional scheme originally conceived for Danish by Diderichsen (1962: 160 ff; see Holmes & Enger 2018: 411 for a recent application to Norwegian), or else within a generative framework (e.g. Nilsen 2000). See Section 4 for the classificatory framework used in the present study.

The last research question of this study concerns lexical priming. This is a theory developed by Hoey (2005) to “account for the existence of corpus linguistic phenomena such as collocation and colligation” (Hoey 2017: 3). The theory assumes that

every word is mentally **primed** for collocational use. As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context.
(Hoey 2005: 8, emphasis in original)

1. Swedish and Norwegian are lexically and syntactically similar enough to be mutually intelligible.

For the present purposes, the most relevant points of Hoey's claims about lexical priming are the following (numbered 1, 4 and 10 in Hoey 2005: 13):

- Every word is primed to occur with particular other words; these are its collocates.
- Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations.
- Every word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations.

In this study of time adverbials, the first two claims concern the phraseological make-up of expressions functioning as time adverbials and the third concerns the placement of individual adverbials. For example, in the phrase *sixty years ago today*, “*sixty* is typically primed for us to occur with *years*, *sixty years ago* is primed to occur as Adjunct, [and] *years* is primed to occur with NUMBER” (Hoey 2017: 19). Hoey also observes that the phrase typically occurs in sentence-initial and possibly text-initial position (2005: 131 f.). A further indication of the importance of the lexical item for the syntactic behaviour of adverbials is offered by Dupont (2019: 244 ff.), who shows that individual contrastive conjuncts and conjunctions may have different lexical primings in both English and French.

It is also important to note Hoey's corollary that all the claims about lexical priming “are in the first place constrained by domain and/or genre” (2005: 13). Words, and combinations of words, “are acquired wrapped in the contexts in which they are encountered” (Hoey 2017: 19). Hence, the primings should not be expected to work in exactly the same way across discourse types; nor across languages, as shown by Dupont (2019), although Hoey (2005, 2017) is concerned with English only.

3. Material and method

The material for this study consists of two collections of news articles from English and Norwegian newspapers. English is represented by the Press reportage section of the BE06 corpus (British English from 2006; see Baker 2009). This corpus was compiled according to the sampling frame of the Brown family of corpora and is thus a one-million-word corpus representing a variety of written genres (Baker 2009: 313). The Norwegian material is a self-compiled collection of news articles from 2011, sampled from online newspapers and the digital archive *Atekst*.² This

2. <<https://www.retriever.no/product/mediemarkiv/>>

material will be referred to as the NoN-corpus, short for ‘Norwegian Newspapers’ but also conceding that it is not a standard corpus. The two corpora are comparable in that both datasets have been sampled from a spread of newspapers (national and regional) and are relatively close to each other in time.

Analysing adverbials, without having predefined the forms to be studied, requires manual excerption (Hasselgård 2010: 6). For this reason, access to complete texts is essential, and the material needs to be limited in size for logistic reasons. The Press reportage category of the BE06 corpus contains 44 files of about 2000 words each (Baker 2009: 317). To make the analysis manageable I have used only the first 22 of these for the present study (texts A01–A22). These represent 17 different newspapers, i.e. five papers are represented with two files.³ The NoN-corpus contains about 23,200 words distributed over nine different newspapers, each contained in one file. The files vary in size from 1545 to 4873 running words (mean = 2582). Each file in both corpora contains more than one news article. Only the newspaper name is given as author in the BE06 documentation and metadata, but it is fair to assume that a number of different writers were involved. As for the Norwegian material, the available bylines reveal that texts may be single- or co-authored, and considering the fact that newspaper articles may be edited by somebody other than the original writer(s), the BE06 practice of listing the newspaper as author seems reasonable. See Table A in the appendix for a full list of the newspapers included, the number of words in each file and the text codes used in the reference tags of the corpus examples.

Since the investigation is based on comparable corpora, the *tertium comparationis*, i.e. the background of sameness for the cross-linguistic comparison (James 1980: 169) must rely on something other than translation evidence (Johansson 2007: 10), namely text comparability and common criteria for identifying the linguistic items under study (Hasselgård 2020: 192). As detailed in the presentation of the corpora, they match in field (news reportage), tenor (newspaper journalists to general public) and mode (written text published online). Furthermore, the time adverbials were identified in both languages on the basis of their temporal meaning and classified according to the same syntactic and functional framework, as explained in Section 4.

4. Classificatory framework for time adverbials

“Temporal adverbials are a rich and heterogeneous category, both in a formal and in a functional respect” (Klein 1994: 147). This section presents the classification

3. The overview of BE06 files was found at <<https://cqpweb.lancs.ac.uk/be2006/>>.

schemes used for the current study in terms of formal realization (4.1), semantic categories (4.2) and syntactic positions (4.3). The classificatory framework comes from Hasselgård (2010), which builds on Quirk et al. (1985) and Biber et al. (1999). It was developed for English, but because of the great similarities in word order between Norwegian and English, it is easily adaptable to Norwegian, as has been demonstrated in previous studies such as Hasselgård (2014a) and Lindquist (1989) for Swedish (see footnote 1). Some temporal adverbials serve mainly as markers of textual organization, or internal cohesion (Martin & Rose 2007: 133), referring to contrasts and continuities between portions of text rather than to the sequencing of events. Potential examples are *then* (see Lewis, present volume), *finally* and *at the same time*. When such adverbials were considered to primarily mark internal cohesion, they were regarded as conjuncts and excluded from the present investigation, which focuses on adjuncts (Hasselgård 2010: 35 f.).

4.1 Syntactic realization of time adjuncts

Time adjuncts have a wide range of syntactic realizations in both English and Norwegian, as listed below. The examples given are equivalent between the languages, unless otherwise stated.

- Single adverb, e.g. *now, already, usually; nå, allerede, vanligvis*
- Adverb phrase, e.g. *so far, earlier today; så langt, tidligere i dag*
- Prepositional phrase (PP), e.g. *in 2005, before midnight; i 2005, før midnatt*. Phrases with *ago* (e.g. *two weeks ago*) are also classified as PPs (Klein 1994: 148). The corresponding Norwegian construction has a split structure as in *for to uker siden* ('for two weeks since').
- Noun phrase (NP), e.g. *last week, this month; forrige uke, denne måneden*
- Finite clause, e.g. *since the troops arrived, before it was closed; siden troppene kom, før det ble stengt*
- Non-finite clause (participle, infinitive, verbless). PPs with a non-finite clause as complement are included in this category (Hasselgård 2010: 37). Examples are *aged four, after taking office, while in a newsagent's*. Norwegian lacks a category that is equivalent to English *-ing* clauses, but can have an infinitive clause after a preposition as in *etter å ha vunnet to valg* ('after to have won two elections').

4.2 Semantic types of time adjuncts

Four types of time adjuncts are recognized in this study. They are listed below along with a brief explanation and one or more examples (see also Hasselgård 2010: 25).

- **Time position** adjuncts refer to a time at which something happened, either a point or a period of time. They are elicited by the probing question *when* (see the italicized part of Example 2).
- **Time duration** adjuncts refer to a period over which something lasted and can mark the beginning, end or total span of that period. They are elicited by the probing question *for how long* (Examples 3-4).
- **Time frequency** adjuncts refer to the number of occurrences or the regularity of an event. They are elicited by the probing question *how often / how many times*, as in the underlined part of (2).
- **Time relationship** adjuncts share many features of time position, but refer to a time that is seen in relation to another (Quirk et al. 1985: 550). Kučera and Trnka (1975: 7) refer to this group as ‘anaphoric time adverbials’. The probing question is *when* (in relation to another time), as in (5).

- (2) POLICE are hunting this man *after a 25-year-old was stabbed four times*... (BE06: A05)
- (3) They remain on the trust's register *until the age of 60*. (BE06: A16)
- (4) And the uncertainty is expected to continue *for weeks* ... (BE06: A02)
- (5) It *finally* succumbed after a failure to sell the Broadford works in the city centre. (BE06: A01)

4.3 Syntactic positions of time adjuncts

The syntactic positions of adjuncts are defined in relation to the verb phrase of the clause (Hasselgård 2010: 40 ff.). This means that only adjuncts belonging to a clause containing a verb have been included in the analysis. In contrast to English, Norwegian is a verb-second (V2) language, which means that some of the positions need to have slightly different definitions in the two languages, as detailed below:

Initial position (I) is before the subject in English clauses except interrogatives. In Norwegian declarative main clauses, initial position is before the finite verb, as in English interrogatives, because the V2 constraint causes inversion of subject and finite when an adverbial occurs clause-initially. See examples (6) and (7).

- (6) *In 2002*, he was sentenced to a total of nine years in prison for two offences. (BE06: A12)
- (7) *På ei uke* er det over 50.000 som har lastet ned appen ... (NoN: KLA)
[‘*In a week* are there over 50,000 who have downloaded the app...’]⁴

4. Norwegian examples are followed by a fairly literal translation into English, enclosed in single quotes. The translations are sometimes deliberately unidiomatic in order to display the

Medial position (M) is after the subject and before a final obligatory constituent. There are three variants of medial position:

- M1: between subject and the verb phrase. In Norwegian this position is only available in dependent clauses and in main clauses with subject-finite inversion, due to the V2 constraint in main declaratives (Examples 8 and 9).
- M2: between an auxiliary and the main verb (Examples 10 and 11).
- M3: between the main verb and an object, complement or obligatory adjunct (Examples 12 and 13).

(8) The couple *now* has five grandchildren. (BE06: A18)

(9) ... at du *allerede* abonnerer på bredbånd. (NoN: VG)
[‘...that you *already* subscribe to broadband.’]

(10) But it has *rarely* been profitable... (BE06: A07)

(11) Forskerne har *lenge* vært enige om at... (NoN: AP)
[‘The researchers have *long* agreed that...’]

(12) Mr Atif said he asked *later* what the hydrogen peroxide was for. (BE06: A21)

(13) Stadig færre nordmenn har *nå* fasttelefon hjemme. (NoN: VG)
[‘Ever fewer Norwegians have *now* landline-phones at home’]

End position (E) is the position after all (other) obligatory constituents; see examples (14) and (15).

(14) The attack happened as the bus travelled past Highgate Underground station in Archway Road. (BE06: A05)

(15) En av tre svarte nei *i 2009*. (NoN: NAT)
[‘One of three answered no in 2009.’]

5. Analysis

5.1 General corpus frequencies

Time adjuncts are numerous in both corpora, thus confirming their importance in the news reportage register. Contrary to expectation, however, the overall frequency of time adjuncts is only marginally higher in Norwegian than in English,

wording of the Norwegian example, while hopefully still making sense to readers who do not read Norwegian.

as shown in Table 1, and the difference is not statistically significant (LL=1.13, $p>0.05$).

Table 1. The frequencies of time adjuncts in the corpora

	Corpus size	Time adjuncts (N)	Per 10,000 words
BE06	43,642	1,100	252.0
Norwegian news (NoN)	23,239	618	265.9

Time adjuncts occur in all the corpus files (see overview in the Appendix), but with varying frequencies. Figure 1 shows a boxplot of the dispersion (normalized per 1,000 words for each file).⁵ It can be observed that both the mean and the median are slightly higher for Norwegian. The lowest value is also higher for Norwegian than for English but the highest values are similar (35 per 1,000 words in English and 34 in Norwegian). However, the smaller interquartile range in Norwegian is most likely due to the smaller number of texts in the NoN-corpus.

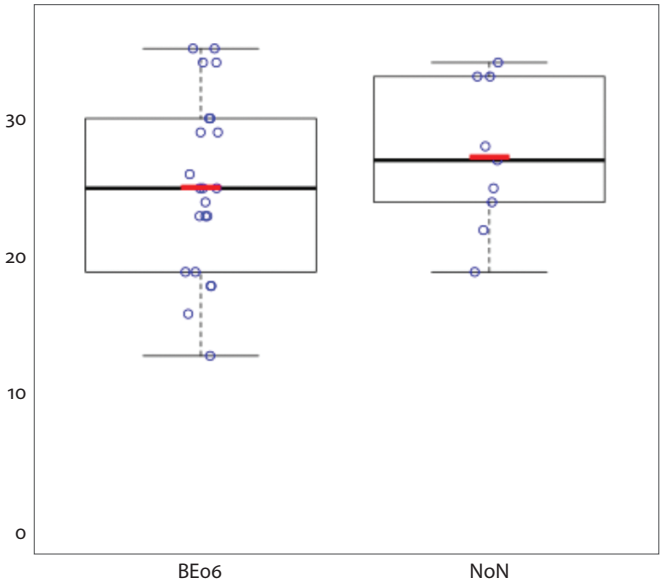


Figure 1. The dispersion of time adjuncts across corpus texts in English and Norwegian

5. The boxplots in this chapter were made with Lancaster Stats Tools Online, Brezina 2018; <<http://corpora.lancs.ac.uk/stats/toolbox.php>>.

5.2 Syntactic and semantic categories of time adjuncts

This section presents the distribution of syntactic and semantic types of time adjuncts in both corpora. Figure 2 shows that the distribution of realization types is very similar between the languages. BE06 has a slightly higher proportion of clauses, particularly non-finite ones, of which the Norwegian material only has three (all infinitive clauses). All the realization types occur in all the files in both languages except non-finite clauses, which occur in 17 out of the 22 BE06 files and three NoN files.

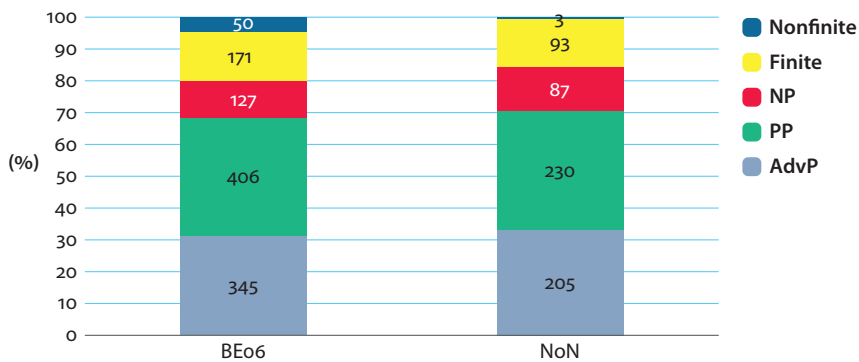


Figure 2. Realizations of time adjuncts in English and Norwegian

The proportions of realization types come out as similar, but it should be noted that equivalent meanings may be realized by different form classes. One difference consists in English adverbs corresponding to Norwegian prepositional phrases (albeit lexicalized ones) in the expressions *yesterday/i går*, *today/i dag*, and *tomorrow/i morgen*;⁶ see examples (16) and (17). Another is references to weekdays, consistently realized by PPs in the English material and by NPs in the Norwegian material, as in (18) and (19); see further Section 5.5.

- (16) BRITAIN is finally falling out of love with the credit card, figures showed
yesterday. (BE06: A21)
- (17) *I går* la de frem 2010-rapporten. (NoN: AP)
[‘Yesterday they presented the 2010 report.’]
- (18) Hungary was plunged into crisis *on Monday* ... (BE06: A22)
- (19) Utplasseringen skjedde *tirsdag*. (NoN: NAT)
[‘The deployment happened Tuesday.’]

6. According to the *OED online* (entry for *tomorrow*), *today* and *tomorrow* also originated as PPs, and could earlier be written as two words or with a hyphen.

In the case of non-finite clauses, English uses two types that have no equivalent in Norwegian, namely *-ing* clauses, which may be augmented by a preposition, as in (20), and past participle clauses introduced by a conjunction as in (21). The only type of non-finite clause functioning as a time adjunct in Norwegian is shown in (22). This construction, with a preposition (*etter* ‘after’) followed by an infinitive, would be ungrammatical in English. It may be an appropriate counterpart of the ‘*after V-ing*’ construction in (20). However, it is infrequent in the NoN corpus, with only three occurrences, whereas the analogous finite construction *etter* + / *that*-clause occurs 14 times.

- (20)

Our pupils did come back early *after finishing one of the exercises they had gone there to do*.

(BE06: A20)
- (21)

When asked about the contracts with private firms, the spokesman added:...

(BE06: A03)
- (22)

Som alternativ har vi et sterkt folkelig mandat *etter å ha vunnet to valg*.

(NoN: KLA)

['As alternative have we a strong popular mandate after to have won two elections.']

The distribution of semantic types of time adjuncts is shown in Table 2. The ranked frequencies of the semantic types are identical across the languages, and their proportions are similar. It is natural that time position is the most frequent adjunct type in news reporting given the communicative focus of this register. All the semantic categories were found in all the corpus files except frequency, which was absent in three BE06 files (A03, A11 and A21).

Table 2. Distribution of semantic types

	BE06		NoN	
	N	%	N	%
Time position	714	64.9	378	61.2
Time duration	132	12.0	70	11.3
Time frequency	71	6.5	44	7.1
Time relationship	183	16.6	126	20.4
	1100	100	618	100

Figure 3 shows the dispersion of the semantic types of time adjuncts across corpus files, with numbers normalized per 2,000 words. The patterns are remarkably similar for position, duration and frequency adjuncts, while relationship adjuncts have a much higher mean frequency as well as greater variation across files in Norwegian than in English.

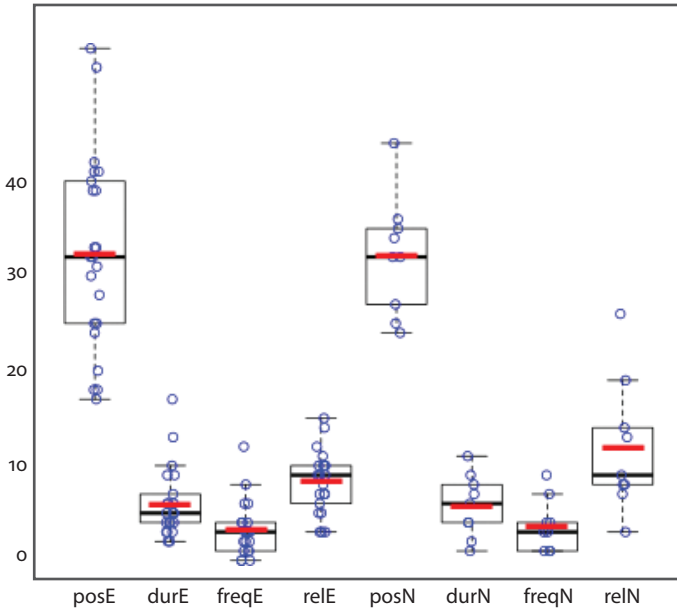


Figure 3. Dispersion of semantic types of time adjunct across corpus files. Normalized frequencies per 2,000 words. English to the left, Norwegian to the right (key to abbreviations: position, duration, frequency, relationship)

5.3 The positions of time adjuncts

Contrastive grammars as well as previous studies of adverbial placement have indicated that adverbial positions differ between English and Scandinavian languages (Norwegian/Swedish) (e.g. Johansson & Lysvåg 1987: 261ff, Lindquist 1989, Altenberg 2010, Hasselgård 2014a). As shown in Figure 4, this is also the case in the present material. The left half of the figure shows the distribution of initial, medial and end position in BE06 calculated as a percentage of the total number of time adjuncts per text, and the right half shows the corresponding distribution in the NoN-corpus. The median and the mean for initial and medial positions are lower in English than in Norwegian, while those for end position are higher. The greater interquartile ranges observed for English are again taken to reflect the greater number of files. A log likelihood test indicates that the aggregate proportions of initial, medial and end position differ significantly across the corpora (LL = 20.08, $p < 0.0001$, Cramer's $V = 0.1084$).⁷

7. The three variants of M were collapsed and cleft position was merged with 'end' in this calculation, performed with the UCREL Significance Test System at <<http://corpora.lancs.ac.uk/sigtest/>>.

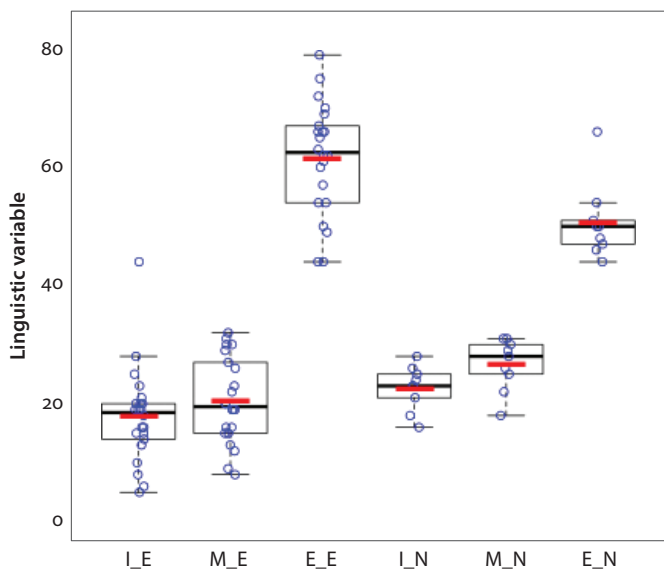


Figure 4. Positions of time adjuncts in BE06 and NoN (percentages of initial, medial and end position per corpus file)

Besides the fact that medial positions are more frequent in Norwegian, the two languages differ as to the type of medial position they prefer, as shown in Table 3. For example, M2 is the most frequent choice in English but the least frequent in Norwegian. It follows from the definition of the medial positions (Section 4) that M2, between an auxiliary and the lexical verb, is available only in complex verb phrases. An examination of the clauses with medial adjuncts reveals that M2 is preferred in clauses with a complex verb phrase (90% in English and 66% in Norwegian). However, in simple verb phrases, British English has a slight preference for M1 over M3 (56% vs. 44%), while Norwegian shows the opposite preference (41% M1 vs. 59% M3).

Table 3. Positions of time adjuncts in BE06 and NoN (raw numbers)

	Initial	M1	M2	M3	E	cleft	
BE06	197	72	91	54	686	0	1100
NoN	138	62	41	60	311	6	618

The cleft position, merged with end position in Figure 4 but shown separately in Table 3, was found only in Norwegian in the present material; see (23). However,

this placement may occur in English too (Hasselgård 2010: 44), as the (idiomatic) translation of (23) demonstrates.

- (23) Jacobsen forteller at det var *først i fjor* at VG mobil ble et eget selskap...
(NoN: KLA).
[‘Jacobsen says that it was only last year that VG mobile became a separate company’]

The various syntactic and semantic categories of time adjuncts have different positional preferences, as shown in Figures 5 and 6. Figure 5 shows a high degree of similarity between the two languages. Adjuncts realized by adverb phrases are the only type to prefer medial positions in both languages; all other realization types are most frequently found in end position. Prepositional phrases differ the most between the languages: for English PPs, medial position is marginal, but it is used for about 20% of the Norwegian PPs, which suggests that medial PPs are less marked in Norwegian than in English. A similar pattern is seen with NPs, though not as clearly (most likely due to the idiosyncratic behaviour of the NP *last night*, as discussed below). Clauses prefer end position in about $\frac{3}{4}$ of the cases in both languages and otherwise tend to go for initial position (see also Hasselgård 2017b).

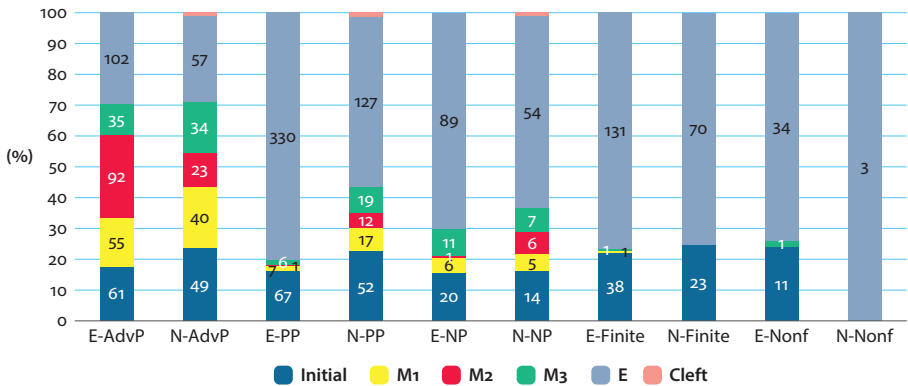


Figure 5. Positions of realization types of time adjuncts

Figure 6 shows that the semantic types differ in their positional preferences in both languages. The figure also reveals some cross-linguistic differences. While the patterns for frequency and relationship adjuncts are relatively similar between the languages, medial positions are used more often for position and duration adjuncts in Norwegian than in English. It may be noted that frequency and relationship adjuncts tend to be realized by adverb phrases, and thus display similar positional patterns to those of AdvPs in Figure 5.

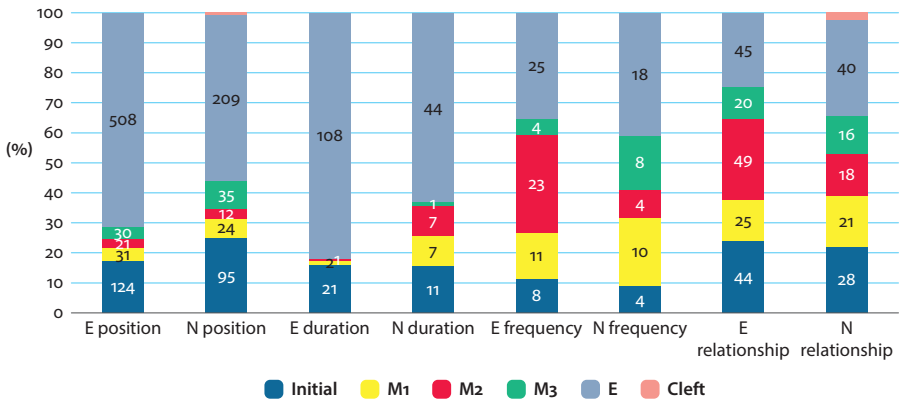


Figure 6. Positions of semantic types of time adjuncts

As noted above, English and Norwegian differ as to the frequency with which prepositional phrases appear in medial position. One reason for this is the frequent use of PPs denoting time position, such as *i går* ('yesterday') and *i morgen* ('tomorrow'); see further next section. However, medial position in Norwegian also accommodates longer PPs rather painlessly, as in (24). This is in contrast to English medial position, which tends to employ some sort of parenthetical marking (Hasselgård 2010: 107 f.) except in M3 position before a heavy post-verbal constituent, as in (25).

- (24) Franske politikere fra både høyre og venstre har *i flere tiår* latt seg invitere på gratis opphold i land som Marokko, Tunisia og Egypt... (NoN: ADR)
 ['French politicians from both right and left have *for many decades* let themselves be invited to free stays in countries like Morocco, Tunisia and Egypt...']
- (25) The YEP reported *in November* how Enid, 31, is being given sanctuary in All Hallows Church in Hyde Park, Leeds... (BE06: A11)

5.4 The lexical realizations of time adjuncts

This section examines lexical realizations of time adjuncts in order to discover patterns of recurring lexis and phraseology. Table 4 shows the 20 most frequent lexical types in both languages. The lexical items in shaded cells have a translational counterpart in the top 20 list in the other corpus. The realization types have been somewhat generalized by representing some of them as colligations (e.g. *when*-clause, *after* NP). Furthermore, numbers have been replaced by # (thus "in ####" represents e.g. "in 2006" and "on ## MONTH" e.g. "on 6 February") and names of weekdays and months by the words WEEKDAY/UKEDAG and MONTH/MÅNED.

Table 4. Lexical realizations of time adjuncts in English and Norwegian (raw numbers (#) and frequencies per 10,000 words)

Rank	BE06			NoN			Gloss
	#	/10k	Item	#	/10k	Item	
1	75	17.2	<i>when</i> -clause	46	19.8	<i>nå</i>	'now'
2	60	13.7	<i>after</i> -clause	32	13.8	<i>da</i> -clause	'when-clause'
3	52	11.9	<i>yesterday</i>	24	10.3	UKEDAG	'WEEKDAY'
4	39	8.9	<i>after</i> NP	22	9.5	<i>etter</i> NP	'after NP'
5	34	7.8	<i>today</i>	22	9.5	<i>når</i> -clause	'when-clause'
6	33	7.6	<i>now</i>	17	7.3	<i>i går</i>	'yesterday'
7	32	7.3	<i>last night</i>	16	6.9	<i>i #####</i>	'in #####'
8	29	6.6	<i>in #####</i>	15	6.5	<i>etter-at</i> -clause	'after (that)-clause'
9	27	6.2	<i>as</i> -clause	12	5.2	<i>da</i>	'then'
10	21	4.8	<i>before</i> -clause	12	5.2	<i>i dag</i>	'today'
11	21	4.8	<i>then</i>	11	4.7	<i>samtidig</i>	'at the same time'
12	19	4.4	<i>still</i>	10	4.3	<i>for...siden</i>	'...ago'
13	18	4.1	<i>never</i>	8	3.4	<i>fortsatt</i>	'still'
14	17	3.9	<i>already</i>	8	3.4	<i>i løpet av</i> NP	'in the course of NP'
15	15	3.4	<i>during</i> NP	8	3.4	<i>tidligere</i>	'earlier'
16	15	3.4	<i>following</i> NP	8	3.4	## MÅNED	'## MONTH'
17	15	3.4	<i>in</i> MONTH	7	3.0	<i>deretter</i>	'thereafter'
18	15	3.4	<i>on</i> WEEKDAY	6	2.6	<i>først</i>	'first'
19	14	3.2	<i>in</i> NP	6	2.6	<i>i</i> MÅNED #####	'in MONTH #####'
20	12	2.7	## <i>years ago</i>	6	2.6	<i>lenge</i>	'for a long time'
	12	2.7	<i>on</i> MONTH ##				
	12	2.7	<i>while</i> -clause				

The degree of recurrence of lexical items (per 10,000 words) is similar between the languages: the top Norwegian item (*nå*) is slightly more frequent than the top English one (*when*-clause), but the frequencies from rank 2 to 20 differ little between the languages. It is noteworthy that the two most frequent English items are clauses while the Norwegian list starts with a simplex. However, Norwegian clauses with *da* and *når* (ranks 2 and 4) both correspond to English *when*-clauses (although they are not interchangeable), and the combined frequency of *da/når* far exceeds that of *when*. *After*-clauses are, by contrast, rather more frequent than their closest Norwegian counterpart, *etter-at*-clauses (rank 8).

More than half of the top-20 items occur in shaded cells, which points to substantial similarity between the languages as regards the composition of news reports. Most of the shared items are realized in English by adverbs (some of which correspond to Norwegian PPs, such as *today/i dag*) or PPs (two of which correspond to Norwegian NPs (*on WEEKDAY* and *on MONTH ##* vs. *WEEKDAY* and *## MONTH*). Semantically, all of the shared top-20 items are time position and relationship adjuncts.

Looking beyond the top 20 lists, we find only a small number of complete recurrent multi-word time expressions in the material. Those that occur more than once per 10,000 words and in at least two corpus files are listed below (in descending order of frequency). Their raw frequencies range from 32 to 5 in English and from 17 to 3 in Norwegian.

- English: *last night, last year, for the first time, at the time, next year, this year, this week, last month, last week*
- Norwegian: *i går, i dag, i fjor, i går kveld, flere ganger, i høst, så langt, hele tiden, denne uken* ['yesterday, today, last year, last night, several times, this autumn, so far, all the time, this week']

The two lists show different collocational and colligational patterns, above all that most of the English items are NPs and most of the Norwegian ones are PPs. This is partly explained by the fact that English has productive patterns with a determiner (*last, this, next*) + a time-noun while Norwegian typically uses the preposition *i* + time-noun in the most recurrent expressions such as *i går* ('yesterday'), *i dag* ('today'), *i fjor* ('last year'). Examples (26) and (27) may serve as illustrations.

(26) *I fjor var det 85 fødsler på Lærdal sjukehus...* (NoN: AP)
 ['Last year were there 85 births at Lærdal hospital...']

(27) *EMI sales fell 10 per cent last year.* (BE06: A13)

5.5 Lexical priming

This part of the investigation concerns cross-linguistic patterns of lexical priming (Hoey 2005) in the case of time adjuncts. To this end, I will examine the collocational, colligational and positional behaviour of the most recurrent lexical realizations of time adjuncts in both languages, represented by the two most frequent types of simplex, phrase and clause. The selection of the most frequent expressions is motivated by the fact that only recurrent expressions are able to display any kind of pattern. However, to expand the empirical base, I also consult other corpora of English and Norwegian speech and writing. Information on these corpora can be found at the end of the References section. Collocation and colligation

are considered only insofar as the phraseological make-up of the time adjuncts is concerned, not with regard to their wider context (see Section 2). The textual collocation of the adjuncts (Hoey 2005: 13) is operationalized here to concern their position in the clause.

Table 5 displays the adjuncts selected for this part of the study along with their raw frequencies and the number of corpus files in which they occur. See Table 4 for their normalized frequencies. The last two lines in Table 5 contain words and phrases which are translation counterparts of the most frequent ones in Norwegian or English, albeit not themselves at the top of the frequency lists. Only one of the items listed in Table 5 occurs in all the relevant corpus files, but the most frequent members of each category all occur in at least half of them.

Table 5. Adjuncts for lexical priming study (# = raw frequencies, NF = number of files out of 22 and 9, respectively)

	English		Norwegian				
		#	NF		Gloss	#	NF
Simplex	<i>yesterday</i>	52	15	<i>nå</i>	'now'	46	9
	<i>today</i>	34	11	<i>UKEDAG</i>	'WEEKDAY'	24	8
Phrase	<i>after NP</i>	43	14	<i>etter NP</i>	'after NP'	22	7
	<i>last night</i>	32	12	<i>i går</i>	'yesterday'	17	6
Clause	<i>when-clause</i>	75	19	<i>da-clause</i>	'when-clause'	32	8
	<i>after-clause</i>	60	19	<i>når-clause</i>	'when-clause'	22	8
Less frequent 'twins'	<i>now</i>	33	18	<i>i dag</i>	'today'	12	9
	<i>on WEEKDAY</i>	15	11	<i>i går kveld</i>	'last night'	5	2

It was noted in Sections 5.2 and 5.4 that time adjuncts with equivalent meanings may be realized by different form classes. For example, two frequent and lexicalised Norwegian PPs, *i går* and *i dag*, correspond to the English single adverbs *yesterday* and *today*. Furthermore, equivalent meanings may be expressed by different phrase types, as shown in Section 5.4 for the NP *last night* corresponding to the PP *i går kveld*, which might be translated literally as 'yesterday evening'. *Last night* might be rendered more congruently in Norwegian as *forrige kveld(en)* or *sist kveld*. These expressions are infrequent in the 1.5-billion-word *Norsk Aviskorpus* ('Norwegian Newspaper Corpus'), with 32 and 9 hits respectively, and absent from the NoN-corpus. In contrast, *i går kveld* occurs more than 75,000 times in *Norsk Aviskorpus*, whereas *yesterday night* occurs four times in the *British National Corpus* (BNC), but not in the BE06, whereas the analogous *yesterday evening* occurs 90 times in the BNC (the majority in 'Fiction and verse') and once in BE06, shown

as (28). The corpus evidence thus indicates that the synonyms *night* and *evening* have different collocational and colligational patterns regarding their co-occurrence with a determiner or an adverb, both of which differ from the preferred Norwegian PP construction *i kveld* ('in evening').

- (28) It was still not under full control late *yesterday evening*. (BE06: A03)

The different ways of constructing weekday adjuncts, with and without an initial preposition, were described in Section 5.4. The observed pattern is that English uses a PP while Norwegian uses a nominal simplex for the equivalent meaning.⁸ In terms of lexical priming, the colligational tendencies of this pair thus differ between English and Norwegian. Interestingly, these patterns may be register-dependent: while all weekday adjuncts in the BE06 news reports were PPs, as in (29), there are examples of the NP use in other registers, see (30) from the spoken part of the BNC, although even in speech the PP form is the most frequent choice in a random sample of 100 occurrences of *Sunday*.

- (29) *On Saturday*, talent scouts will attend the graduate show of the fashion college Central Saint Martins... (BE06: A08)

- (30) and you could go *Sunday* and have a beer... (BNC-spoken)

In the NoN-corpus, weekday adjuncts are typically NPs, as in (31), except for one example of PP realization, shown in (32). However, in a Norwegian corpus of informal conversation (NoTa), 24 out of a total of 29 time position adjuncts involving the weekday *søndag* ('Sunday') were PPs, as exemplified in (33).⁹ It is thus possible that the collocational and colligational patterns of weekday adjuncts are register-dependent in both (British) English and Norwegian, but that the principles of priming work differently in the two languages.

- (31) Når jeg er på ferie, er jeg ikke utenriksminister, sa hun *lørdag* til fransk radio. (NoN: ADR)

['When I am on holiday, am I not foreign minister, said she *Saturday* to French radio.']

- (32) Det første flyet som er planlagt til Norge, går *på tirsdag* ... (NoN: VG)
['The first flight that is planned for Norway, leaves *on Tuesday*...']

8. As shown in Table 4 (ranks 20 and 16), references to dates are also PPs in English (e.g. *on 15 March*) and NPs in Norwegian (e.g. *15. mars*). However, the (non-)use of the preposition seems less variable with dates than with weekdays in both languages.

9. All occurrences of *søndag* 'Sunday' in NoTa were examined, but those that were not time position adjuncts in the form of a PP or a bare NP were ignored (e.g. *forrige søndag* 'last Sunday' and *fra fredag til søndag* 'from Friday to Sunday').

- (33) men jeg så en veldig fin film på tv *på søndag* (NoTa)
 ['but I saw a really good film on tv *on Sunday*']

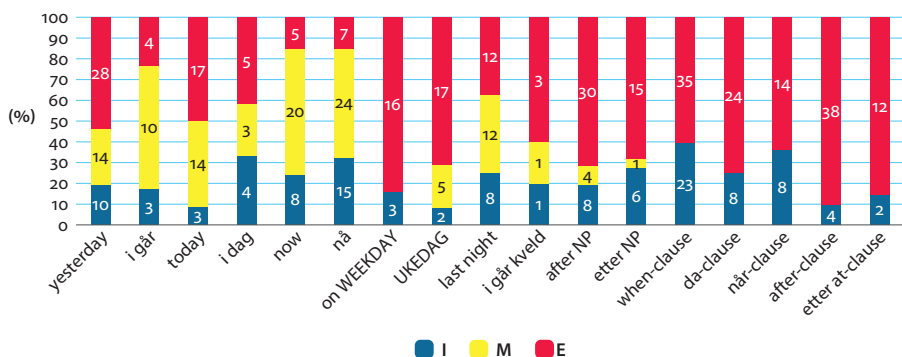


Figure 7. Positional patterns of pairs of lexicogrammatical items

Figure 7 shows the individual patterns of placement of the adjuncts listed in Table 5, i.e. the two most frequent simplexes, phrases and clauses in either language, juxtaposed with their closest translation counterpart in the other language. That is, it compares translationally related pairs of lexicogrammatical items with regard to positional priming. The patterns can be compared to those shown in Figure 5 for different realization types of adjuncts. The pairings in Figure 7 are obviously complicated by the fact that the members of several pairs belong to different form classes. But the intra-lingual comparison also shows idiosyncratic behaviour of members of the same form class, e.g. the adverbs *yesterday*, *today*, and *now*. The former two are relatively similar, but *now* has a very different pattern with medial position being the most frequent and end position the least frequent choice. As regards *yesterday* and *today*, the former seems more easily primed to initial position and slightly less easily to medial position than the latter. This may be connected to phonological factors as well as the expectedness and information value of time references in the news reports register.

Cross-linguistically we note that the cognates *now* and *nå* behave very similarly with regard to position. *Yesterday* and *i går* have rather different positional patterns, in that *yesterday* selects end position more than 50% of the time while the most common position of *i går* is medial position. The analogous pair *today* and *i dag* have slightly more similar positional patterns, but there are too few examples in the material to draw reliable conclusions. This is even more clearly the case for the phrase *i går kveld*, which was included as a counterpart of *last night*. However, with only five occurrences in two corpus files it can be concluded that *i går kveld* is not primed to occur as an adjunct in this register to the same extent as the English

last night. The English *yesterday* and *today* use medial position less than the overall tendency for English adverb phrases (cf. Figure 5), which may be related to their function as position adjuncts rather than frequency or relationship adjuncts (cf. Figure 6). Conversely, the positional pattern of *i går* (to a greater extent than that of *i dag*), is closer to that of Norwegian adverb phrases than to that of prepositional phrases in Figure 5. This is presumably associated with the high degree of lexicalization of these phrases: they are conceptualized as single words.

In both languages the two most frequent phrases have quite distinct patterns of placement, as shown in Figure 7. While *after* NP behaves much like English prepositional phrases in general, the pattern of *last night* is more like adverbs (Figure 5), suggesting that this phrase too is conceptualized much as a single word. Like its English counterpart *after* NP, *etter* NP has a relatively similar pattern to that of Norwegian PPs in general, albeit with a lower proportion of medial position, which may be due to the small numbers involved. Cross-linguistically, Figure 7 shows that the patterns of *etter/after* NP are strikingly similar.

As regards the positional patterns of the clauses in Figure 7, it is notable that they occur only in initial and end position in both English and Norwegian in the present material. Norwegian *da*-clauses follow the overall pattern of Norwegian finite clauses (cf. Figure 5), i.e. initial position is selected in just over one in five cases. *Når*-clauses select initial position more often, almost 40%, like English *when*-clauses. This is in line with previous studies of *when*-clauses, e.g. Ford (1993) and Hasselgård (2017b). Both *after*-clauses and the much less frequent Norwegian counterpart with *etter at* select initial position much less frequently than *when/når*-clauses (c. 10% vs. c. 40%) and also less frequently than the average for finite clauses in both languages (Figure 5).

To summarize this section, we have seen a collocational difference between the languages in that the words *next* and *last* collocate with a time noun more consistently in English to produce time expressions (often paralleled by PPs in Norwegian), while syntactically similar Norwegian NPs are less common. Furthermore, there is a colligational difference between the languages in that names of weekdays are primed to occur without a preposition in the news register in Norwegian but not in (British) English. Textually, this investigation has shown that frequent deictic (short) adjuncts seem primed to occur in non-final position more often than other adjuncts (*nå/now*, *i går*, *last night*), and furthermore, that the more lexicalized a phrase is, the more it deviates in positional preference compared to phrases of the same type (as was the case with e.g. *i går*, *i dag*, *last night*). It seems likely that the primings of especially *i går*, *i dag*, *last night*, *yesterday* and *today* are specific to the register of news reports, in which texts are typically read the day after they are written, or even on the same day, in the case of online publication.

6. Summary of findings and concluding remarks

The present chapter has investigated English and Norwegian time adjuncts in news reports. The overall impression is that there is great similarity between the two languages in this area. First, the overall frequencies of time adjuncts were unexpectedly similar, in view of previous studies of fictional language in the two languages. Furthermore, the distribution of syntactic realizations and semantic types of adjuncts also turned out to be rather similar, with no significant differences between the two datasets. However, it was noted that non-finite clauses are marginal as time adjuncts in Norwegian, but more widespread in English.

Cross-linguistic differences concern first and foremost the placement of time adjuncts. While there is some similarity in this area too, particularly the fact that end position is the most frequent one for most syntactic and semantic types of time adjuncts, it was shown that Norwegian makes more use of medial position. While medial position in English tends to allow only very short adverbials, preferably those realized by single adverbs (Hasselgård 2010: 290), Norwegian medial position more easily accommodates both NPs and PPs without any indication that they are parenthetical. Hence there are more position and duration adjuncts in medial position in Norwegian than in English, where this position is to a large extent reserved for frequency and relationship adjuncts and a small number of short time position adjuncts with atypical primings (Section 5.5).

The study of lexical realizations showed a high degree of similarity between the languages in that more than half of the 20 most frequent lexical items were equivalent between the two lists. However, to arrive at the frequency lists, some of the lexical items were abstracted into colligations such as *after* NP, *in* ####, and *when*-clause. Relatively few of the multi-word lexical items were highly recurrent in identical form in either English or Norwegian. There was little overlap between the recurrent multi-word lexical items, highlighting the fact that the same meanings can be realized by different form types in the two languages. In particular, some Norwegian PPs correspond to English adverbs and NPs, for example *i går* vs. *yesterday* and *i fjor* vs. *last year*. Some of the productive NP patterns in English are also available in Norwegian, such as *neste/forrige* ('next'/'last') + time noun. However, these patterns are less consistently used in Norwegian, particularly *forrige*, which felicitously collocates with the nouns *uke* ('week') and *måned* ('month'), but not with nouns corresponding to *night* and *year*, which both frequently collocate with *last* in English. A larger material is needed to ascertain whether the differences uncovered are due mainly to differences in the realization of certain meanings, or if English and Norwegian newspapers choose to talk about time in different ways.

Probing further into the lexical primings of the most recurrent time adjuncts in both languages, it was found that Norwegian collocations with the preposition

i ('in'), e.g. *i går*, *i morgen*) are not matched by English PPs, but by adverbs and NPs. Conversely, references to weekdays consistently involved a preposition in English but not in Norwegian. However, as both the PP and the NP patterns exist in both languages, other corpora containing other registers were consulted, and the findings suggest that weekday adjuncts may have different colligational primings for PP vs. NP realization across registers in both languages. The investigation also revealed individual patterns of colligational primings (including placement) across the most frequent lexical realizations. Deictic adjuncts with 'expected information', such as *last night*, *yesterday*, *today* and their Norwegian counterparts, seem primed to occur in non-final positions in this register, due to its special temporal context, with reading taking place very shortly after publication and texts typically reporting on very recent events.

Methodologically, the manual bottom-up approach used in this study is necessary for (close to) full recall of time adjuncts unless the investigation is limited to a set of predefined expressions. However, it severely limits the size of the dataset, as manual excerption is time-consuming. A potential follow-up might be to target specific expressions emerging from this study as particularly interesting because of either their frequency or their cross-linguistic differences and investigate these on the basis of more material.

The attention to lexical realizations and lexical priming has not been common in contrastive studies of adverbials (an exception being Dupont 2019), but I believe it can give new insights into language-specific and register-specific patterns of time adjuncts. In particular, it highlights the interrelationship between lexis and grammar by revealing specific patterns of individual lexical realizations against the backdrop of the general patterns of their grammatical class. In spite of the small size of the material and low degree of recurrence, the present study was able to point to some cross-linguistic similarities and differences in lexical priming. Again, larger corpora would be needed to get more reliable data on the lexical priming of time adjuncts. Another interesting avenue of further research would be to extend the comparison to more registers in both languages.

Johansson (2012: 64) argues that the contrastive analysis can throw "special features of the languages compared into relief, including preferred ways of expressing similar meanings". This study has taken a bottom-up approach to the expression of (similar) temporal meanings in English and Norwegian by means of adjunct adverbials. If we assume with Hoey (2005: 14) that corpus frequencies can give indications of lexical priming, the corpus approach applied here should be able to provide insights into what is natural, not only what is possible, in the use of time adjuncts (ibid.: 2) in both Norwegian and English. In a cross-linguistic perspective this might act as a peephole into the idiomaticity of both of the languages compared.

References

- Altenberg, Bengt. 2010. Conclusive English *then* and Swedish *då*: A corpus-based contrastive study. *Languages in Contrast* 10(1): 102–123. <https://doi.org/10.1075/lic.10.1.05alt>
- Baker, Paul. 2009. The BE06 Corpus of British English and recent language change. *International Journal of Corpus Linguistics* 14(3): 312–337. <https://doi.org/10.1075/ijcl.14.3.02bak>
- Biber, Douglas & Conrad, Susan. 2009. *Register, Genre and Style*. Cambridge: CUP.
<https://doi.org/10.1017/CBO9780511814358>
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics. A Practical Guide*. Cambridge: CUP.
<https://doi.org/10.1017/9781316410899>
- Diderichsen, Paul. 1962. *Elementær Dansk Grammatik*. 3. udgave. København: Gyldendal.
- Dupont, Maïté. 2019. Conjunctive Markers of Contrast in English and French. From Syntax to Lexis and Discourse. PhD dissertation, University of Louvain. <<https://dial.uclouvain.be/pr/boreal/object/boreal:215212>> (1 April 2021).
- Ebeling, Signe Oksefjell & Ebeling, Jarle. 2017. A cross-linguistic comparison of recurrent word combinations in a comparable corpus of English and Norwegian fiction. In *Contrasting English and Other Languages through Corpora*, Markéta Janebová, Ekaterina Lapshinova-Koltunski & Michaela Martinková (eds), 2–31. Newcastle upon Tyne: Cambridge Scholars.
- Faarlund, Jan Terje, Vannebo, Kjell Ivar & Lie, Svein. 1997. *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.
- Ford, Cecilia E. 1993. *Grammar in Interaction. Adverbial Clauses in American English Conversations*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511554278>
- Hasselgård, Hilde. 2010. *Adjunct Adverbials in English*. Cambridge: CUP.
<https://doi.org/10.1017/CBO9780511676253>
- Hasselgård, Hilde. 2014a. Discourse-structuring functions of initial adverbials in English and Norwegian news and fiction. In *Genre- and Register-related Discourse Features in Contrast*, Marie-Aude Lefer & Svetlana Vogeleer (eds). Special issue of *Languages in Contrast* 14(1): 73–92.
- Hasselgård, Hilde. 2014b. Conditional clauses in English and Norwegian. In *Affaire(s) de grammair. Mélanges offerts à Marianne Hobæk Haff à l'occasion de ses soixante-cinq ans*, Hans-Petter Helland & Christine M. Salvesen (eds), 185–204. Oslo: Novus.
- Hasselgård, Hilde. 2017a. Temporal expressions in English and Norwegian. In *Contrasting English and Other Languages through Corpora*, Markéta Janebová, Ekaterina Lapshinova-Koltunski & Michaela Martinková (eds), 75–101. Newcastle upon Tyne: Cambridge Scholars.
- Hasselgård, Hilde. 2017b. Adverbial clauses in English and Norwegian fiction and news. In *Contrastive Analysis of Discourse-pragmatic Aspects of Linguistic Genres* [Yearbook of Corpus Linguistics and Pragmatics], Karin Aijmer & Diana Lewis (eds), 119–139. Cham: Springer. https://doi.org/10.1007/978-3-319-54556-1_6
- Hasselgård, Hilde. 2020. Corpus-based contrastive studies: Beginnings, developments and directions. *Languages in Contrast* 20(2): 184–208. <https://doi.org/10.1075/lic.00015.has>
- Hoey, Michael. 2005. *Lexical Priming. A New Theory of Words and Language*. London: Routledge.
- Hoey, Michael. 2017. Cohesion and coherence in a content-specific corpus. In *Lexical Priming. Applications and Advances*, Michael Pace-Sigge & Katie J. Patterson (eds), 3–40. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.79.01hoe>

- Holmes, Philip & Enger, Hans-Olav. 2018. *Norwegian. A Comprehensive Grammar*. Oxon: Routledge.
- James, Carl. 1980. *Contrastive Analysis*. London: Longman.
- Johansson, Stig. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies* [Studies in Corpus Linguistics 26]. Amsterdam: John Benjamins.
<https://doi.org/10.1075/scl.26>
- Johansson, Stig. 2012. Cross-linguistic perspectives. In *English Corpus Linguistics: Crossing Paths*, Merja Kytö (ed.), 45–68. Amsterdam: Rodopi.
- Johansson, Stig & Lysvåg, Per. 1987. *Understanding English Grammar, Part II: A Closer View*. Oslo: Universitetsforlaget.
- Klein, Wolfgang. 1994. *Time in Language*. London: Routledge.
- Kučera, Henry & Trnka, Karla. 1975. *Time in Language: Temporal Adverbial Constructions in Czech, Russian, and English*. Ann Arbor MI: Michigan Slavic Publications.
- Lindquist, Hans. 1989. *English Adverbials in Translation. A Corpus Study of Swedish Renderings*. Lund: Lund University Press.
- Martin, J. R. & Rose, David. 2007. *Working with Discourse. Meaning beyond the Clause*. London: Continuum.
- Nilsen, Øystein. 2000. *The Syntax of Circumstantial Adverbials*. Oslo: Novus Press.
- OED Online. Oxford: OUP <<https://oed.com>> (1 December 2019).
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Virtanen, Tuija. 2014. Sentence-initial adverbials in written texts: On discourse functions and cognitive motivations. In *Adverbials in Use. From Predicative to Discourse Functions*, Laure Sarda, Shirley Carter-Thomas, Benjamin Fagard & Michel Charolles (eds), 103–133. Louvain: Presses Universitaires de Louvain.

Corpora used

- BE06, *British English 06*. <<http://www.helsinki.fi/varieng/CoRD/corpora/BE06/>>, Baker (2009)
- BNC, British National Corpus. <<http://www.natcorp.ox.ac.uk/>>
- NoN, Norwegian newspapers: See description in Section 3.
- Norsk Aviskorpus (Norwegian Newspaper Corpus). <<https://www.nb.no/sprakbanken/show?serial=oai:clarino.uib.no:avis-plain&lang=>>>
- NoTa, *Norsk Talemålskorpus* (Norwegian Speech Corpus). <<http://www.tekstlab.uio.no/nota/oslo/english.html>>

Appendix

Table A. Details of the corpus composition

English newspapers	code	words	Norwegian newspapers	code	words
Aberdeen Evening Post	A01	2007	Adresseavisen	ADR	1932
Daily Mail	A02	1813	Aftenposten	AP	4873
Daily Mail	A03	2045	Dagsavisen	DAV	3600
Daily Star	A04	1996	Dagens Næringsliv	DN	1545
Evening Standard	A05	1928	Klassekampen	KLA	2968
The Guardian	A06	1937	Nationen	NAT	1685
The Independent	A07	1975	Stavanger Aftenblad	STA	2417
The Independent	A08	1900	VG	VG	2356
The Sun	A09	2012	Vårt Land	VL	1863
The Sun	A10	2078			
Yorkshire Evening Post	A11	2031			
Western Morning News (Plymouth)	A12	1958			
Yorkshire Evening Post	A13	1972			
Times Higher Education Supplement	A14	1944			
South Wales Echo	A15	2031			
The Sentinel (Stoke)	A16	1992			
The Scotsman	A17	1939			
Nottingham Evening Post	A18	1946			
The Northern Echo	A19	1950			
Birmingham Evening Mail	A20	2080			
Daily Telegraph	A21	2105			
Daily Telegraph	A22	2003			
		43642			23239

Minutes of action! A contrastive analysis of time expressions in English and Norwegian football match reports

Signe Oksefjell Ebeling

University of Oslo

This contrastive study examines the use of two time expressions – *on # minutes* and *etter # minutes* ('after # minutes') – in football match reports with the aim of shedding light on their conditions of use in English and Norwegian. The cross-linguistic findings suggest that the patterns are typically used to report on the achievements of players. With regard to syntax, the English pattern is clearly preferred in final position, while the Norwegian pattern is found in either initial or final position, with a slight preference for initial position. A more inconclusive difference can be noted regarding the patterns' distribution across the 90 minutes; the English pattern is more evenly distributed across the two halves of the match than the Norwegian one.

1. Introduction

Previous corpus-based contrastive research involving English and Norwegian has largely focused on the language of fiction, often on the basis of the *English-Norwegian Parallel Corpus*.¹ Less has been done on other text types, although several studies have emerged in recent years, such as Fløttum et al. (2006) on research articles in English, French and Norwegian, Hasselgård (2014), which includes English-Norwegian comparable newspaper data, and Rørvik and Monsen (2018) with comparable data from English and Norwegian research articles in the field of didactics. To further contribute to broadening the contrastive perspective with other text types, this paper investigates contrastive data from *The English-Norwegian*

1. See e.g. the ENPC bibliography <https://www.hf.uio.no/ilos/tjenester/kunnskap/sprak/omc/enpc_omc_publications_2021update.pdf>

Match Report Corpus (ENMaRC). The ENMaRC is a comparable corpus consisting of written football match reports from two seasons in the English Premier League and the Norwegian Eliteserie (see further Section 3.1).

Previous contrastive studies of English and Norwegian fiction texts have shown that time adverbials are frequent in both languages (Hasselgård 2014; Ebeling & Ebeling 2017). It has also been shown that time expressions abound in other text types, notably (oral) sports commentaries (Hasselgård 2010) and football match reports (Ebeling 2019).

Drawing on findings from the abovementioned studies, and in the context of the theme of this volume – time in languages, this study examines a set of time expressions used in online football match reports in a contrastive perspective with the aim of shedding light on their conditions of use in English vs. Norwegian. The starting point of the investigation is the most frequently occurring lexical words referring to time in a comparable corpus of English and Norwegian match reports, namely *minutes* and *minutter* ('minutes'). By pinpointing the lexico-grammatical and phraseological characteristics of these items, the study seeks to contribute to a better understanding of how time expressions are used to frame events in football match reports. Moreover, a detailed investigation of these patterns may also contribute to more comprehensive cross-linguistic knowledge of the use of time expressions in English and Norwegian in general.

Although initial observations of *minutes* and *minutter* in the material drawn from the ENMaRC show substantial overlap between the two languages in terms of phraseological patterning, there are also some differences worth noting. English has more different recurrent patterns with *minutes* (see Section 4), i.e. there is less variation in the Norwegian data and some patterns are unique to English. To narrow the scope, the current study focuses on the predominant pattern in each language (see Sections 4.1 and 4.2), i.e. the English sequence *on # minutes* (1), and the Norwegian sequence *etter # minutter* ('after # minutes') (2).

- (1) ... Alexis smashed home a crucial third goal *on 83 minutes*. (AFC)²
- (2) *Etter 86 minutter* fikk vi likevel en god sjanse ... (AaFK)
[After 86 minutes got we even so a good chance']

Further observations of a subset of the data suggest that there may be interesting cross-linguistic differences in the use of the expressions and way in which events at a specific moment in time are reported, leading to the following research questions:

-
2. The identifiers in brackets after each example refer to the football clubs, e.g. AFC = Arsenal Football Club. For a full overview of all club abbreviations, see the Appendix.

- a. what is the preferred placement of the time expression in the clause;
- b. which participant is most prominently taking part in the action when the time expression is used, and to what extent does the action refer to the most important part of the game, namely goal scoring;
- c. when, during the 90 minutes of a game, is the time expression typically used?

With regard to question (a), previous contrastive research suggests that there may be a general difference in the placement of adjunct adverbials (including time adjuncts) between English and Norwegian (e.g. Hasselgård 2004, 2014). Norwegian has been found to more easily accept such adverbials in initial position than English, although final position seems to be preferred in both languages (Hasselgård 2014). However, a more nuanced and complex picture seems to emerge as more registers are investigated and a finer grained semantic analysis of individual adverbials are taken into consideration (see e.g. Hasselgård this volume). It will thus be interesting to see to what extent the more general trend applies to the text type studied here. Also in connection with research question (a), which is lexicogrammatical in nature, a subsidiary question of tense use will be considered, as this may have an impact on how the action is viewed.

The other questions – (b) and (c) – are narrower in scope than (a) in being less general and more specifically tied to the match report text type. To my knowledge, no previous study has addressed these two issues in a contrastive perspective.

The paper is organized as follows: Section 2 outlines some relevant previous contrastive research, while Section 3 is devoted to a description of the material and method used. Section 4 starts out with an overview of the patterns with *minutes/ minutter* identified in the corpus, before focusing on the most frequent ‘minute’ pattern in each language, addressing the research questions in turn. Section 4.4 includes a discussion of the main cross-linguistic findings regarding the two patterns, and finally, some concluding remarks are offered in Section 5.

2. Previous research

The language of sports in general, and of football in particular, has received some attention in (corpus) linguistic research, notably in the edited volumes by Lavic et al. (2008) and Callies and Levin (2019). The current paper is largely inspired by the results from two studies from these volumes: Levin (2008) and Ebeling (2019).

Drawing on material from newspapers, Levin finds that the language of English football match reports consists to a large extent of recurrent sequences of words with conventionalized functions to describe events. A telling example, illustrating this pervasive use of pre-constructed phrases in newspaper match reports,

is given in Levin's example from the *British National Corpus* (3), which "consists entirely of holistically stored, slightly variable items (TEAM NAME, *take the lead, in the nth minute and against the run of play*)" (Levin 2008: 151).

- (3) Villa took the lead in the 15th minute against the run of play. (BNC)

Most relevant in the current context is what Levin refers to as the "phraseology of football time". Indeed, newspaper match reports seem to have developed their own register-specific phraseology to specify time. Moreover, Levin notes that "[f]ootball time is often specified in minutes" (2008: 150), and in his data the sequence *in the nth minute* is among the most common ones used to express time. Expressions with the plural *minutes* also abound, notably *the dying minutes, the opening minutes, and minutes from time*.

Ebeling (2019) is a contrastive study using material from a smaller version of the *English-Norwegian Match Report Corpus* than the one used in the present study. In the 2019 study, several corpus linguistic methods were explored to investigate some characteristics of the language of football match reports in English and Norwegian. Different types of frequency lists were extracted and scrutinised, pointing to some salient features emerging from word lists, n-gram lists and keyword lists in this highly specialised comparable corpus.

The n-gram lists, consisting of 3- and 4-grams, showed a high frequency, and great overlap, of time expressions in English and Norwegian, including (*in*) *the nth minute* / *i det (x) minutt, the first half* / *i første omgang* ('in first half'). While both the English and Norwegian match reporters are keen to report on when the action takes place (and the Norwegian ones more so than the English), the English reporters are also concerned with where on the pitch it happens, e.g. the 3-grams *the edge of*, as shown in example (4).

- (4) Okazaki went for goal from *the edge of* the penalty area ... (MFC)

These observations are more or less in line with findings from previous contrastive studies of English and Norwegian fiction texts. For example, both Ebeling et al. (2013) and Hasselgård (2017) note that temporal expressions are more frequently used in Norwegian fiction than in English fiction, while Ebeling and Ebeling (2017) found that spatial expressions seem to be more frequently used in English fiction when compared to Norwegian fiction. The fact that these two text types (fiction and match reports) seem to behave similarly in their use of temporal and spatial expressions may point to some underlying systemic or cultural differences between English and Norwegian.

With regard to the first research question, it should be noted that Hasselgård (2004), drawing on English and Norwegian fiction data, found that Norwegian uses adjuncts, including time adjuncts, in initial position more often than English

does, although the preferred position in both languages seems to be final position. In fact, adjuncts in initial position are regarded as marked themes in English (cf. Hasselgård 2014, with reference to Halliday & Matthiessen 2004). She further observes that there are register differences in the use of adjuncts in English (cf. Hasselgård 2010). Thus, we may expect to find the Norwegian minute expression more often in initial position than the English one, although the text type may play a role in this respect.

Finally, both Levin (2008) and Ebeling (2019) note that frequently occurring sequences typically refer to goals (not) scored. This is commonly done by sequences including *net/nettet* ('the net') and *goal/mål*. A case in point is Levin's example of *hit the ball in the net* = goal and *have the ball in the net* = disallowed goal. While these phraseologies may not seem to be directly relevant to the present study focusing on *minutes* and *minutter*, there is reason to believe that a time reference is often included in connection with the event of (not) scoring a goal, as shown in example (5).

- (5) ... and when he did *have the ball in the back of the net* three *minutes* after the restart, the flag was up for offside. (MU)

3. Material and method

3.1 The corpus

The *English-Norwegian Match Report Corpus* (ENMaRC) is a comparable corpus of football match reports published online by the football clubs themselves. The ENMaRC contains reports from two seasons in the English Premier League (2016–17 and 2017–18) and the Norwegian Eliteserie (2017 and 2018).³ The reason for compiling such a corpus in the first place was two-fold: a personal interest in (the language of) football and the fact that the language of football, and indeed sports in general, has been shown to be a fascinating text type from a linguistic point of view, e.g. in terms of idiomaticity, phraseology, metaphor and lexico-grammar (see, e.g. the edited volumes by Lavric et al. 2008 and Callies & Levin 2019 and papers by Berg & Ohlander, e.g. 2012, 2016). To add a contrastive dimension on top of this seemed like a good avenue for further broadening our knowledge of cross-linguistic/cross-register differences and similarities between languages.

3. The football season in England runs from August until May (hence 2016–17 and 2017–18), while the Norwegian season runs from March until November (hence 2017 and 2018).

The Premier League (PL) sub-corpus contains match reports from 23 different teams, while the Eliteserie (ES) sub-corpus contains reports from 18 teams.⁴ Some teams are represented with reports from both seasons, others from only one as they were either relegated from, or promoted to, the PL/ES in the 2017–18 / 2018 season. Each season, there are 20 teams in the PL and 16 in the ES; thus, the maximum number of reports per team per season in the PL is 38 and in the ES 30. Moreover, some of the teams did not publish reports after every match, while others only published in one of the seasons covered. The difference in number of teams, as well as a notable difference in how long the reports are, has resulted in two sub-corpora of quite different sizes: the PL corpus contains roughly 990,000 words and the ES corpus contains less than a third of that, with roughly 315,000 words. For an overview of the clubs and corpus texts, see Tables A and B in the Appendix. A more detailed description of the compilation process of the ENMaRC can be found in Ebeling (2019).

3.2 The method

Being a contrastive study drawing on comparable data, the current investigation applies the following criteria of comparability to ensure a relatively sound *tertium comparationis*, i.e. “background of sameness against which [cross-linguistic] differences are to be measured” (Ringbom 1994: 738).

- text type = match reports;
- date of publication = from the same years/seasons;
- topic = football at the highest competitive level in two countries;
- situational context = written by in-house journalists primarily for fans of the respective clubs;
- object of study = most frequently occurring temporal sequences containing the cognates *minutes*/*minutter*.

Against this background of sameness, it is likely that the cross-linguistic observations will be comparable and valid.

In a bottom-up approach, AntConc (Anthony 2019) was used to extract wordlists from the English and Norwegian sub-corpora of the ENMaRC. The lists drew my attention to the plural cognate nouns *minutes* and *minutter*, which are ranked 18 in both the English and the Norwegian data, with 5,553 and 2,385

4. The sub-corpora contain 23 and 18 files, respectively, each holding all the match reports for one club.

occurrences, respectively.⁵ Interestingly, *minutes* is in fact the highest ranked lexical word in the English match reports, while *minutter* is only preceded by *ballen* ('the ball') at rank 14 in the Norwegian ones. Thus, two seemingly identical time-referring nouns were found to be extremely frequent in the material and therefore thought to reflect the importance of situating the action on the pitch in time.⁶ In other words, they are arguably the single-most important time-framing device in these football match reports, and as such interesting to investigate further.

4. Overview and analysis of 'minute' patterns

It was noted above that *minutes* and *minutter* are among the most frequent words in the English and Norwegian football match reports. One explanation for this lies in the very nature of the game of football which is confined to two halves of 45 minutes each; however, there may also be a linguistic explanation for their frequent use. Sinclair (1999) and Sinclair et al. (1970/2004) argue that frequent words are frequent because they occur in frequent phrases, or "frequent phrasal constructions which express conventional pragmatic functions in text" (Stubbs 2007: Section 5; see also Summers 1996: 263). Indeed, Levin (2008) finds that the language of English football match reports to a large extent consists of recurrent sequences of words with conventionalized functions to describe events. Thus, potentially recurrent word-combinations that form more or less fixed phrases with *minutes* and *minutter* may be a contributing factor to their frequent use.

After scrutinising the use of *minutes* and *minutter* in context, using different display techniques in AntConc,⁷ I found that Sinclair's hypothesis seems to hold. In the English match reports, more than 4,500 of the 5,553 occurrences of *minutes* occur in so-called "frequent phrases", or patterns, defined here as recurring 3–4-word sequences of which *minutes/minutter* is part. The most salient sequences with *minutes*, listed according to frequency, include:

5. The datasets have a mean of 5.8 vs. 7.1 per 1,000 s-units, respectively. The difference is not statistically significant (t-test: $p < 0.1$; Wilcoxon: $p < 0.1$; tests performed in R; R version 3.6.2).

6. For comparison, *minutes* is found at rank 546 in the BNC (using the search engine Phrases in English, <http://phrasesinenglish.org/>), with 17,598 occurrences. The normalized frequency in the BNC overall is 20 per 100,000 words, whereas the English match reports reach a staggering 560 per 100,000 words. The text type in the BNC that has the highest attested number of *minutes* is Newspaper with 36 occurrences per 100,000 words.

7. Concordance lines sorting on -3L – +3R and Clusters/N-grams with *minutes/minutter* as the left-most/right-most word.

- *on* # *minutes* (> 1,100 occ.)
- *with* (*just*) # *minutes* V (> 600 occ.)
- (*just*) (*/#) *minutes later* (> 600 occ.)
- *after* (*) # *minutes* (> 300 occ.)
- # *minutes before* (NP) (> 250 occ.)
- # *minutes after* (NP) (> 250 occ.)
- # *minutes into* NP (> 150 occ.)
- # *minutes from time* (> 150 occ.)
- *with* (*just*) # *minutes on the clock* (> 100 occ.)

As noted in the introduction, the single-most frequent combination with *minutes* in the data is *on* # *minutes* with 1,114 occurrences, while the sequences *with* (*just*) # *minutes* V and (*just*) (*/#) *minutes later* occur more than 600 times each in slightly different variants; in Erman and Warren's (2000: 40) terms they are "extendible" in that they allow optional elements. The most frequent English pattern is only extendible insofar as the # represents different numbers. Examples (6) and (7) show two typical variants of the former pattern and (8) and (9) two variants of the latter.

- (6) Puel made his second change *with 16 minutes remaining*, ... (SFC)
- (7) ... had the ball in the net *with just three minutes played* ... (MFC)
- (8) *Five minutes later*, City doubled their advantage. (SFC)
- (9) Palace needed a quick response, and thankfully they got it
just a couple of minutes later ... (CPFC)

Similarly, Norwegian *minutter* is found in recurrent patterns in more than 1,900 out of the 2,385 occurrences. The most frequent ones include:

- *etter* (*) # *minutter* ('after' (*) # *minutes*') (> 850 occ.)
- # *minutter senere* ('# *minutes later*') (> 350 occ.)
- # *minutter før* NP ('# *minutes before* NP') (> 250 occ.)
- # *minutter etter* NP ('# *minutes after* NP') (> 100 occ.)
- # *minutter ut i* NP ('# *minutes into* (lit.: out in) NP') (> 100 occ.)

Etter (*) # *minutter* is by far the most frequent recurrent sequence in the Norwegian data; although it is extendible, it most often occurs with no intervening item between the preposition *etter* ('after') and # *minutter*, as in example (10). However, some extensions can be noted, with adverbs such as *bare* and *kun* ('only/just'), as in (11).

- (10) *Etter 50 minutter* må Kalludra byttes ut på grunn av en skade. (KFK)
[After 50 minutes Kalludra has to be replaced because of an injury']

- (11) ... og SF var i ledelsen *etter kun tre minutter*. (SF)
 [... and SF were in the lead after just three minutes']

It is worth noting that, although both Levin (2008) and Ebeling (2019) comment on the frequent use of *MINUTE/MINUTT*, neither comment on the plural forms of the lemmas in the most frequently attested patterns in the current material. This may have to do with the methods used for extracting and identifying patterns or, in the case of Levin, the material investigated – in his case newspapers.

There are some notable differences between the phraseological patterning in the English and Norwegian match reports. First, the lists above are only intended to give a snapshot of the patterns, with a cut-off point of roughly 100 occurrences, and are therefore not exhaustive. The general tendency, taking all identified patterns into account, is for English *minutes* to enter into more different patterns than Norwegian *minutter*: between 15 and 20 for *minutes* and between 7 and 9 for *minutter*, depending on the exact cut-off point for something to count as a pattern.⁸ This discrepancy in the number of patterns suggests that there is less variation in the potential for patterning with *minutter* compared to *minutes*. While this could possibly be attributed to the higher overall frequency of *minutes*, I would argue that, with the number of occurrences attested for *minutter*, most possible patterns would have revealed themselves in the amount of data at hand.

Second, comparing the lists of the most frequent patterns, we can note that all the Norwegian ones have formally similar and intuitively semantically equivalent or corresponding patterns in English, e.g. *etter # minutter* \approx *after # minutes*; *# minutter senere* \approx *# minutes later*. Further, two of the nine most frequent patterns with *minutes* seem to be unique to English: *on # minutes* and *# minutes from time*, whereas *'with (just) # minutes V'* and *with (just) # minutes on the clock* have similar realizations in Norwegian that do not reach the threshold of 100 occurrences: *'med # minutter igjen V'* ('with # minutes left V') and *med # minutter igjen på klokka* ('with # minutes left on the clock'). It is hard to determine why two of the English patterns do not have direct counterparts in Norwegian, and we can only conclude that they are lexicalized in English and not in Norwegian in this corpus. Thus it has to do with idiomaticity and preferred ways of saying things.

However, as noted, both languages have a clearly preferred 'minute' pattern – *on # minutes* and *etter (*) # minutter*, accounting for 20% and 36% of all occurrences of *minutes* and *minutter*, respectively. These patterns will be the focus of the contrastive case study described below in order to establish to what extent they are employed similarly or differently when referring to the action on the pitch.

8. An exact cut-off point has not been set for the purpose of this study, but to qualify as a pattern the 3-4-grams have to recur in identical form and be used by several clubs.

4.1 Preferred patterns in English and Norwegian: Position in clause and tense

This section gives a more detailed account of the *on # minutes* and *etter # minutter* patterns, focusing on the first of the three points of potential cross-linguistic interest mentioned in the introduction, namely (a) placement of the pattern in the clause and the tense it combines with.

First, the number of occurrences of the two patterns in initial, medial and final position was registered. Initial is defined as before the subject, medial as between the subject and main verb, and final as following the main verb and other obligatory syntactic elements.⁹ Examples (12)-(17) illustrate the three positions of *on # minutes* and *etter # minutter* in the material.

- (12) *On 75 minutes* Bolasie combined with Lukaku before shooting over ... (AFCB)
- (13) ... a neat Hughes pass *on 14 minutes* found Richarlison near the left edge of ... (TH)
- (14) ... Bournemouth went close to doubling their lead *on 24 minutes*. (HC)
- (15) *Etter 24 minutter* var det tid for Samuel igjen. (VFK)
[‘After 24 minutes it was time for Samuel again’]
- (16) ... da Sandefjord *etter 56 minutter* for første gang i kampen gikk opp i ledelsen. (STB)
[‘... when Sandefjord after 56 minutes for the first time in the match took the lead’]
- (17) Brann får en stor sjanse *etter 17 minutter*, ... (TIL)
[‘Brann gets a big chance after 17 minutes’]

As shown in Figure 1, the tendency for the English *on # minutes* pattern is quite clear: final position is by far the preferred position. It is similar, in this respect, to the ‘on weekday’ adverbial in newspapers, analysed by Hasselgård (this volume; see her Figure 5).

While the distribution of instances for final position ranges from 0–80 per 1,000 s-units (orthographic sentence) across the 23 PL files, with a median of almost 20, the median for initial is just above 1, and medial position is really marginal, with a median of 0.6.

The picture for the Norwegian *etter # minutter* pattern is quite different, as shown in Figure 2.

9. See Hasselgård, Section 4.1 this volume, for a detailed description of the three positions in the two languages.

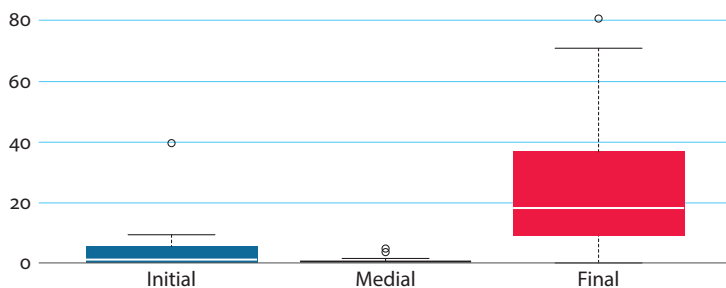


Figure 1. Position in clause of *on # minutes* per 1,000 s-units

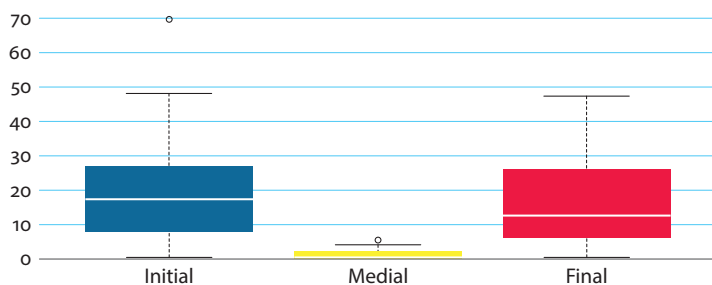


Figure 2. Distribution of *etter # minutter* according to position in the clause

In terms of position in the clause, *etter # minutter* seems to go as easily in initial as in end position. The plot in Figure 2 only suggests a slightly higher median for initial position per 1,000 s-units (17.4 for initial and 12.5 for final), and the difference is not statistically significant between initial and final (Wilcoxon: $p = 0.35$; $r = 0.2$).¹⁰

Although the match reports are written immediately after the games are finished, it is relevant, and potentially interesting, from a cross-linguistic point of view to compare the use of tense with the most frequent ‘minute’ patterns.¹¹ Will the past time setting of the reports automatically trigger the past tense or do the journalists resort to the historic present to report on past events? The historic present is conventionally used “to make the narrative appear more vivid by assimilating it to the here-and-now of the speech act” (Huddleston & Pullum 2002: 130). With this in mind, the tense was also recorded as part of the lexico-grammatical

10. As there are only 18 data points in the material, these measures should be interpreted with caution.

11. Thanks to Mats Johansson for suggesting this.

description of the two main ‘minute’ patterns, as this may further shed light on the patterns’ conditions of use.

It can be noted that the past tense is overwhelmingly used in *on # minutes* contexts, as in examples (12) to (14). In fact, only one instance of the present tense was found in sentences containing the pattern, which incidentally is found in a list of key points in one of the match reports (18).

- (18) - Sturridge opens scoring *on 35 minutes* from magnificent Coutinho pass. (LFC)

The remaining 1,113 sentences with *on # minutes* contain a simple past tense form in the main clause (19) or, in the surrounding sentences, in the case of modal verb phrases (20).

- (19) This, however, was always going to be Pablo Zabaleta’s night, with the Argentine cult hero coming on to a standing ovation *on 62 minutes* ... (MC)
- (20) ... a fierce strike punched wide by Lukasz Fabianski, and then forced the keeper to stretch and gather a dangerous low centre at the second attempt. The Pole could do nothing about our opener *on 18 minutes*, however. His defence failed to fully clear a testing Branislav Ivanovic cross, ... (CFC)

Tense choice with *etter # minutter* is slightly different; although the past tense is clearly favoured (with 666 occurrences), as in examples (15)-(17), the present tense also features (195 occurrences), as shown in example (21).¹²

- (21) *Etter 75 minutter* kommer også den etterlengte [sic] scoringen. (OBK)
[‘After 75 minutes comes also the long-awaited goal’]

However, it is important to note that as many as 77 of the 195 instances of the present tense are found in the reports from one club, namely Kristiansund (KBK). Nevertheless, the present tense has been used by 11 out of the 18 clubs, suggesting that it is not a feature that is exclusive to one individual team or journalist.

4.2 Preferred patterns in English and Norwegian: Participants

When it comes to research question (b), on the use of the *on # minutes* and *etter # minutter* patterns and the kinds of participants that are most prominent in the action

12. The remaining 14 sentences that contain *etter # minutter* do not contain a verb (e.g. *Kjem-pemulighet etter 32 minutter*. (STB) ‘Great chance after 32 minutes!’); the surrounding sentences contain a verb in the present tense in two cases and the past tense in 12 cases.

reported, these are typically (part of) the subject and include Player, Event, Home team, Away team and Other. Examples of each category is given in (22) to (31):

Participants:

– Player

(22) Sterling was again unlucky *on 55 minutes* ... (EFC)

(23) *Etter 36 minutter* var Erik Brenden nær ved å sette inn utlikningen. (LSK)
[‘After 36 minutes Erik Brenden was close to score the equalizer’]

– Event (typically an inanimate entity in contexts with no clear/explicit animate (human) participant involved)

(24) The winner arrived *on 88 minutes* in fortunate fashion ... (EFC)
(winner = ‘match-winning goal’)

(25) Forsøket i tverrliggeren *etter 25 minutter* var den største. (OBK)
[‘The attempt in the crossbar after 25 minutes was the biggest’]

– Home team (team writing the report)

(26) The Foxes were in front *on six minutes* ... (LC)

(27) Men *etter 20 minutter* tok LSK likevel ledelsen. (LSK)
[‘But after 20 minutes LSK took the lead even so’]

– Away team (the opposition of the team writing the report)

(28) Everton were level *on 77 minutes* ... (AFCB)

(29) Aalesund var farlig frempå *etter 38 minutter* ... (VIF)
[‘Aalesund were dangerously advancing after 38 minutes’]

– Other (spectators, manager, referee)

(30) They were out of their seats *on 75 minutes* ... (WFC)

(31) *Etter tolv minutter* trodde de fremmøtte at hjemmelaget hadde tatt ledelsen. (S08)

[‘After twelve minutes those present thought that the home team had taken the lead’]

From Figures 3 and 4 below it becomes quite clear that the players are in focus and referred to when something happens *on* or *etter* (‘after’) a given minute of the game in both English and Norwegian. Events, home team and away team are given more or less equal attention in the match reports, while other participants are only occasionally referred to.

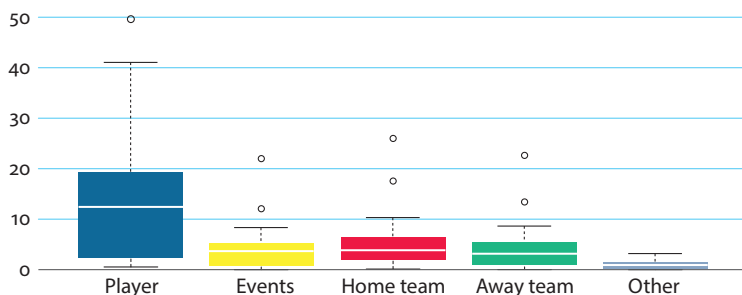


Figure 3. Participant involved in contexts with *on # minutes* per 1,000 s-units

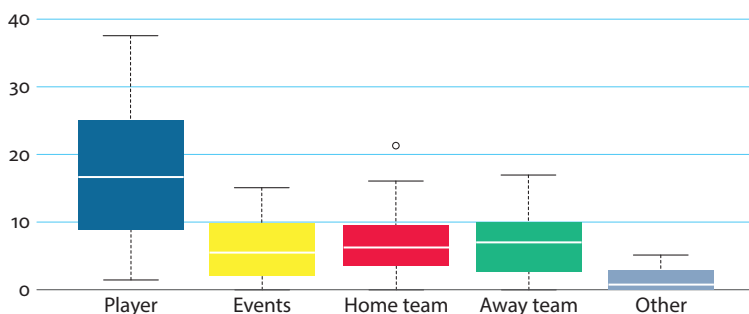


Figure 4. Participants with *etter # minutter* per 1,000 s-units

4.3 The preferred patterns in English and Norwegian: When are the expressions used?

Since the minute is clearly of the essence in football, it is relevant to produce a minute-by-minute timeline for the patterns to establish if there are specific minutes that are more prominent than others, i.e. which of the 90 minutes is most often referred to in the match reports. Thus, with reference to research question (c) above, when, during the 90 minutes of a game, is the pattern used? Figure 5 outlines the trend for the *on # minutes* pattern.

On # minutes seems to be used to report equally frequently on the action in the two halves, peaking at certain intervals: around the 13th and 27th minute in the first half and the 55th, 65th, 70th and 77th minute in the second half. It is also worth noting that the pattern is not used beyond the 90-minute mark.

When it comes to the Norwegian pattern, Figure 6 shows that the use of *etter # minutter* peaks in the beginning of the games, as well as after around 20 minutes. It does not feature as prominently in the second half, but has a slight peak around

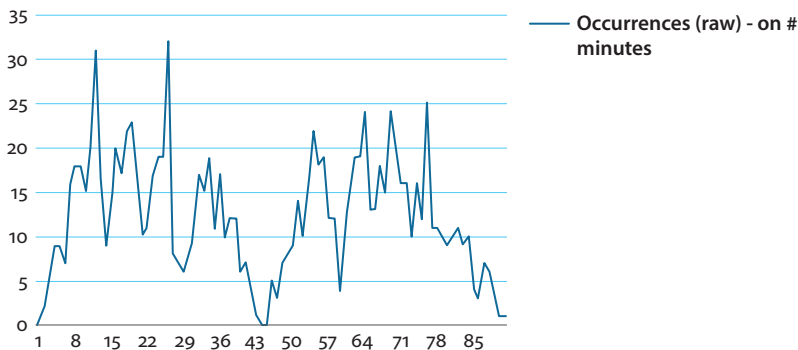


Figure 5. Occurrences of *on # minutes* distributed across 90+ minutes

the 65th minute, and, although not frequently used beyond 90 minutes, there are a handful of examples where this is the case, e.g. *etter 92 minutter*.

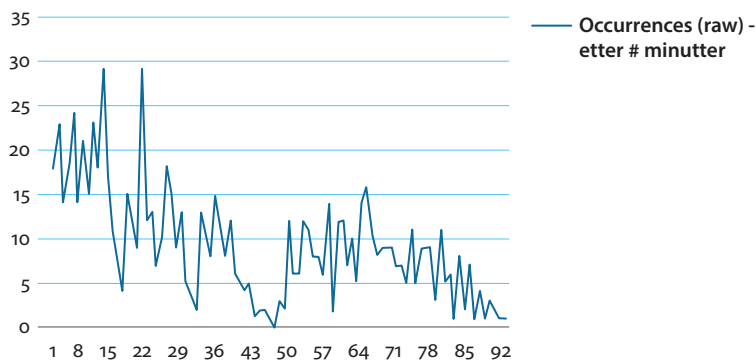


Figure 6. Occurrences of *etter # minutter* distributed across 90+ minutes

As a natural extension of the use of these expressions during a game, it is of interest to explore the extent to which the *on # minutes* and *etter # minutter* patterns are used in connection with the most important thing in football, namely to score goals (see the brief discussion on this towards the end of Section 2). In the Premier League material, *on # minutes* is indeed used in connection with scoring goals, as in example (32), but more often it is used in contexts referring to other events (665 vs. 449 occurrences, i.e. 60% of the cases), as in examples (33) and (34).

(32) Everton added a third on 79 minutes... (MC)

(33) We almost added a third on 68 minutes... (TH)

(34) Lys Mousset was introduced from the bench on 58 minutes... (AFCB)

In other words, the action taking place on a specific minute most often refers to some other aspect that is considered important to report on, e.g. missed opportunity (33), substitution (34), etc. Admittedly, example (33) is concerned with the goal-scoring part of the game, but as a goal was not actually scored, examples like this were considered as part of the “goal-not-scored” category.

In the Eliteserie material, *etter # minutter* is also used to report both on goals scored and other events taking place on the pitch, as in examples (35) and (36), respectively. However, there is a clear tendency to use the pattern to refer to events that do not involve a goal. In fact, this is the case in 72% of the occurrences (633 vs. 242).

- (35) Stølås' scoring *etter 55 minutter* ble matchavgjørende ... (VIF)
 ['Stølås' goal after 55 minutes was match decisive ...']
- (36) En hands i feltet førte til straffespark *etter 32 minutter* ... (RBK)
 ['A handball in the area led to a penalty after 32 minutes ...']

4.4 Discussion of the preferred patterns in a contrastive perspective

The most commonly used ‘minute’ pattern in Norwegian is proportionally more frequent than the most common pattern in English, with 38 vs. 30 occurrences per 1,000 s-units, respectively. This is in line with previous cross-linguistic findings that report on the use of time adverbials in English vs. Norwegian (Ebeling et al. 2013; Hasselgård 2017; see also Hasselgård, this volume, Section 5.1).¹³ Nevertheless, the two patterns clearly represent the preferred phraseology of football (minute) time in their respective languages, and are as such comparable.

It is also interesting to note that the most frequent patterns in the two languages differ in one important aspect, namely the prepositions: *on* vs. *etter* (‘after’). The *Oxford English Dictionary*’s entry for *on* includes the definition “[o]f time, or action implying time [...] [i]ndicating the day or part of the day when an event takes place; = [...] on the instant”. In other words the English pattern is arguably more focused on a point in time than the most frequent Norwegian pattern, where *etter* suggests that the focus is more generally to report on a past event, i.e. something that follows a completed event in the past (cf. *etter* 1.1.4, *Det Norske Akademis Ordbok*). However, in the current context, *on* seems to be close in meaning to *etter*, but further study is needed to establish this with more certainty.

In the introduction, three potential cross-linguistic points of difference between the two languages were put forward on the basis of preliminary

13. However, according to a Wilcoxon rank sum test, the difference is not statistically significant ($p = 0.1$).

observations of the data, forming the basis for the research questions. Drawing on the full data set it has become clear that the main contrastive difference in the use of the two patterns applies to the first point, namely position in the clause. While the English pattern is overwhelmingly preferred in final position, the Norwegian pattern is more evenly distributed between initial and final, being marginally preferred in initial position. The difference in use of initial position of the English and Norwegian expressions is visualised in Figure 7, while the lack of a marked difference in the use of final position is shown in Figure 8.

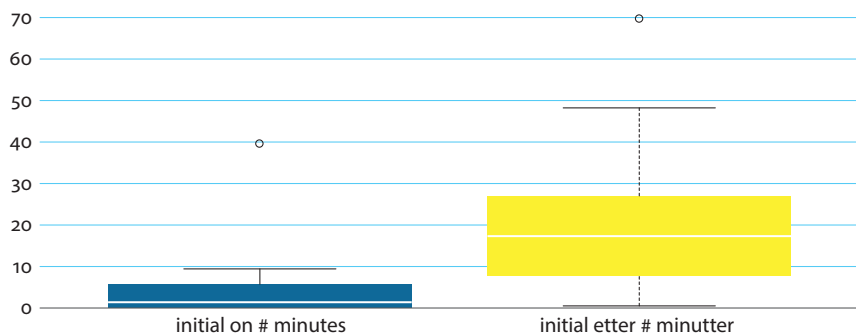


Figure 7. Distribution of initial *on # minutes* vs. initial *etter # minutter* per 1,000 s-units

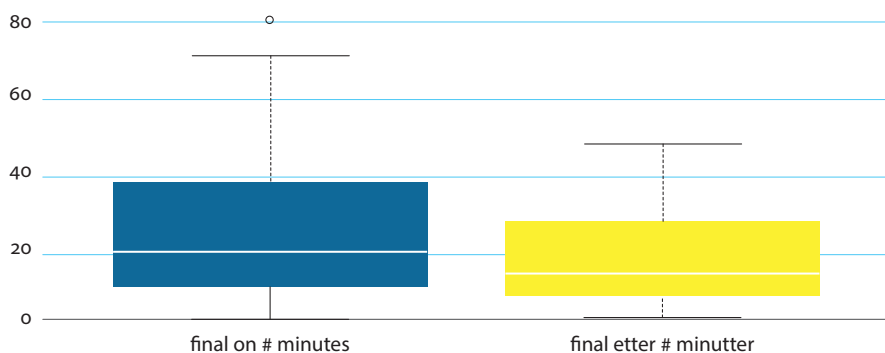


Figure 8. Distribution of final *on # minutes* vs. final *etter # minutter* per 1,000 s-units

The finding regarding the preferred positions of these time expressions in a highly specialised text type reflects results from previous contrastive studies of English and Norwegian, where it has been noted that adjuncts occur more naturally in initial position in Norwegian than in English (e.g. Hasselgård 2014). A possible reason for this could be the tendency in Norwegian to use initial adjuncts for discourse-linking purposes (ibid.: 74). However, the fact that the *etter # minutter*

pattern occurs slightly more often in initial position than in final position in the Norwegian match reports suggests that this text type deviates from the norm, at least in Norwegian. Thus, while the current study contributes further evidence to there being a difference between the two languages when it comes to (time) adverbial placement, this difference is even more pronounced in the match reports for these two 'minute' patterns, and as such this text type is even more prone to using such patterns as a linking (and structuring) device in Norwegian.

Some minor differences were noted with regard to tense use with the patterns; English exclusively uses the past tense with *on # minutes*, whereas there was some more variation in Norwegian, where both the past and present were represented with *etter # minutter*. As (live) sports commentaries are often set in the present tense, one might have expected that the match reports in the ENMaRC, being written immediately after, or even during, the games, would resort to the immediacy of the historic present. While this does not seem to be the case in the English match reports, could it be that this text type is considered to be more involved (/oral) in nature in Norwegian, and as a result the historic present is more frequently used to report on past events?

Contrary to expectations based on a smaller data set, there were no marked differences in participant prominence in sentences in which the two 'minute' patterns occur. The journalists most often report on what the players do in both languages. Other perspectives, such as events, home and away team were less, but similarly used in the two languages.

Regarding the question of which minutes the patterns typically refer to, some differences between *on # minutes* and *etter # minutter* were noted. The English pattern is more evenly distributed across the 90 minutes of a game than the Norwegian pattern. There is no reason to believe that events worth reporting on are more evenly distributed in the PL compared to the ES. At this stage we can only speculate whether there may be a functional difference between the two patterns or whether Norwegian makes use of other time expressions in the second half. To determine this with any certainty lies outside the scope of the current investigation.

Finally, both 'minute' patterns are more often used in connection with events that do not involve goals being scored, and more so with the Norwegian pattern (72% of the cases) than the English one (60%). The difference between the Norwegian and English pattern could be related to the fact that the Norwegian one is potentially more general than the English one, in the sense that the latter may be considered more dramatic as it (grammatically) refers to a point in time rather than a finished past event, which is the case for *etter # minutter*. It could also be the case that the English pattern is more specialized in the sense that it is more tied to a text type reporting on football or, more broadly, sports in general. An investigation of this claim has to await further study, though.

Following the above analysis and discussion, the typical use of the English *on # minutes* pattern can be summarized in the following way:

- Final position
- Past tense
- Player in focus
- Relatively even distribution across the 90 minutes, with noticeable peaks in each half
- Goal not scored

And similarly, for the Norwegian *etter # minutter* pattern, the uses can be summarized as follows:

- Initial or final position
- Past (and present) tense
- Player in focus
- More frequently used, and with more peaks, in the first half
- Goal not scored

To illustrate these cross-linguistic tendencies, a typical example with *on # minutes* is given in (37) (player participant, *past tense*, *no goal*, *final position* and *second half*) and with *etter # minutter* in (38) (player participant, *past tense*, *no goal*, *initial position* and *first half*).

(37) *Jesus blazed wide when clean through **on 64 minutes**, ...* (TH)

(38) *Etter 12 minutter fikk Lehne Olsen en mulighet etter et godt innlegg fra venstre kant, men headingen gikk over mål.*
 [‘After 12 minutes got Lehne Olsen a chance after a good cross from the left winger but the header went over goal’]

5. Concluding remarks

Drawing on material from a comparable corpus of online football match reports, this contrastive study started out by identifying the most frequent time-referring items in the English and Norwegian reports. The word forms *minutes* and *minutter* were selected for further scrutiny, and it was suggested that they feature so prominently in match reports because they are often used as part of frequent recurring phrases.

Thus, the current study lends support to Levin’s (2008: 150) observation with regard to the phraseology of football time; it is often specified in minutes. This phenomenon does not only apply to English, but seems to be valid for other languages

as well, as attested by the frequent use of both English *minutes* and Norwegian *minutter* as part of recurring phrases in football match reports.

The conditions of use of the two most frequent ‘minute’ patterns in the two languages – *on # minutes* and *etter # minutter* – were found to be very similar. Both were typically found to be used in contexts where the event was seen from a player perspective and did not involve a goal scored. The main difference was found to be syntactic in nature, with the English pattern typically occurring in final position and the Norwegian pattern being divided between initial and final position with a slight preference for initial. This finding is not in full accordance with previous cross-linguistic studies of adverbial placement in English and Norwegian, but strengthens the view that there is a contrastive difference between the two languages in this regard, possibly due to text type and the actual time expressions under investigation. It would be interesting, in a future study, to check to what extent there is a correlation between type of participant and placement of the time expressions in the clause.

The study also uncovered some differences regarding tense use with the ‘minute’ patterns. Although the past tense is the favoured choice in both languages, the Norwegian pattern more readily seems to combine with the historic present, which arguably reports events more vividly (Huddleston & Pullum 2002: 130).

A more inconclusive difference was noted when it came to the patterns’ distribution across the 90 minutes, where the English pattern was found to be more evenly distributed across the two halves than the Norwegian one. A more in-depth study is needed to find out whether the distribution reflects other (functional) differences between the two patterns.

Another avenue for further study would be to take a closer look the English pattern that is formally congruent to the Norwegian *etter # minutter*, namely *after # minutes*. As the contrastive focus in this study was on the most frequent ‘minute’ patterns in the two languages, such a comparison had to be left out. However, we can already conclude that the two intuitively similar patterns differ with regard to how often they are used, which is an interesting starting point for a future contrastive study of the two.

Beyond the concrete contrastive findings regarding the use of ‘minute’ patterns in English and Norwegian, this study has demonstrated the potential of a carefully constructed comparable corpus for use in contrastive analysis between languages. Investigating a hitherto underexplored text type in an English-Norwegian context, the study has contributed further insights into previously held cross-linguistic views at a more general linguistic level.

Acknowledgements

I would like to thank the two anonymous reviewers and the editors of this volume for their constructive comments and suggestions on an earlier version of this chapter.

References

- Anthony, Laurence. 2019. AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University, <<http://www.laurenceanthony.net/software>> (12 August 2019).
- Bergh, Gunnar & Ohlander, Sölve. 2012. Free kicks, dribblers and WAGs. Exploring the language of “the people’s game”. *Moderna språk* 106(1): 11–46. <<http://ojs.ub.gu.se/ojs/index.php/modernasprak/article/view/117>> (13 March 2020).
- Bergh, Gunnar & Ohlander, Sölve. 2016. Iniesta passed and Messi finished clinically: Football verbs and transitivity. *Nordic Journal of English Studies* 15(2): 19–38. <https://doi.org/10.35360/njes.359>
- Callies, Marcus & Levin, Magnus (eds). 2019. *Corpus Approaches to the Language of Sports: Texts, Media, Modalities*. London: Bloomsbury.
- Det Norske Akademis Ordbok (Norwegian Academy Dictionary). 2020. Oslo: Det Norske kadem i for Språk og Litteratur. <<https://www.naob.no/ordbok>> (9 March 2020).
- Ebeling, Jarle, Ebeling, Signe Oksefjell & Hasselgård, Hilde. 2013. Using recurrent word-combinations to explore cross-linguistic differences. In *Advances in Corpus-based Contrastive Linguistics: Studies in Honour of Stig Johansson* [Studies in Corpus Linguistics 54], Karin Aijmer & Bengt Altenberg (eds), 177–199. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.54.11ebe>
- Ebeling, Signe Oksefjell. 2019. The language of football match reports in a contrastive perspective. In *Corpus Approaches to the Language of Sports: Texts, Media, Modalities*, Marcus Callies & Magnus Levin (eds), 37–62. London: Bloomsbury. <https://doi.org/10.5040/9781350088238.ch-003>
- Ebeling, Signe Oksefjell & Ebeling, Jarle. 2017. A cross-linguistic comparison of recurrent word combinations in a comparable corpus of English and Norwegian fiction. In *Contrasting English and Other Languages through Corpora*, Markéta Janebová, Ekaterina Lapshinova-Koltunski & Michaela Martinková (eds), 2–31. Newcastle upon Tyne: Cambridge Scholars.
- Erman, Brit & Warren, Beatrice. 2000. The idiom principle and the open choice principle. *Text* 20: 29–62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Fløttum, Kjersti, Dahl, Trine & Kinn, Torodd. 2006. *Academic Voices. Across Languages and Disciplines*. [Pragmatics and Beyond New Series 148]. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.148>
- Hasselgård, Hilde. 2004. Thematic choice in English and Norwegian. *Functions of Language* 11(2): 187–212. <https://doi.org/10.1075/fol.11.2.03has>
- Hasselgård, Hilde. 2010. *Adjunct Adverbials in English*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511676253>
- Hasselgård, Hilde. 2014. Discourse-structuring functions of initial adverbials in English and Norwegian news and fiction. *Languages in Contrast* 14(1): 73–92. <https://doi.org/10.1075/lic.14.1.05has>
- Hasselgård, Hilde. 2017. Temporal expressions in English and Norwegian. In *Contrasting English and Other Languages through Corpora*, Markéta Janebová, Ekaterina Lapshinova-Koltunski & Michaela Martinková (eds), 75–99. Newcastle upon Tyne: Cambridge Scholars.
- Huddleston, Rodney & Pullum, Geoffrey K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP. <https://doi.org/10.1017/9781316423530>
- Lavric, Eva, Pisek, Gerhard, Skinner, Andrew & Stadler, Wolfgang (eds). 2008. *The Linguistics of Football*. Tübingen: Günter Narr.

- Levin, Magnus. 2008. Hitting the back of the net just before the final whistle: High-frequency phrases in football reporting. In *The Linguistics of Football*, Eva Lavric, Gerhard, Pisek, Andrew Skinner & Wolfgang Stadler (eds), 143–155. Tübingen: Gunter Narr.
- Oxford English Dictionary (OED)*, Online. 2020. Oxford: OUP. <<http://oed.com/>> (10 March 2020).
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>> (20 December 2019).
- Ringbom, Håkan. 1994. Contrastive analysis. In *Encyclopedia of Language and Linguistics* 2, R. E. Asher & J. M. Y. Simpson (eds), 737–742. Oxford: Pergamon.
- Rørvik, Sylvi & Monsen, Marte. 2018. Marked themes in English and Norwegian academic texts in the field of didactics. *Bergen Language and Linguistics Studies* 9(1): 43–68.
<https://doi.org/10.15845/bells.v9i1.1521>
- Sinclair, John. 1999. A way with common words. In *Out of Corpora: Studies in Honour of Stig Johansson* [Studies in Corpus Linguistics 57], Hilde Hasselgård & Signe Oksefjell (eds), 157–179. Amsterdam: Rodopi.
- Sinclair, John, Jones, Susan & Daley, Robert. 1970/2004. *English Collocation Studies: The OSTI Report*. London: Continuum.
- Stubbs, Michael. 2007. Notes on the history of corpus linguistics and empirical semantics. In *Collocations and Idioms*, Marja Nenonen & Sinikka Niemi (eds), 317–329. Joensuu: Joensuu Yliopisto.
- Summers, Della. 1996. Computer lexicography: The importance of representativeness in relation to frequency. In *Using Corpora for Language Research. Studies in Honour of Geoffrey Leech*, Jenny Thomas & Mick Short (eds), 260–266. London: Longman.

Appendix

Table A. Overview of Premiership teams, URLs and s-units in the individual corpus files (setting: hide tags only)¹⁴

Corpus id. / Club abbr. ¹⁴	Team	URL	# of reports	# of s-units (AntConc counts)	# of tokens (AntConc counts)	# of types (AntConc counts)
AFC	Arsenal	http://www.arsenal.com/	76	2,465	49,594	4,117
AFCB	Bournemouth	http://www.afcb.co.uk/	76	2,015	51,700	3,782
BHA	Brighton & Hove Albion	https://www.brightonandhovealbion.com/	38	817	24,067	2,312
BFC	Burnley	http://www.burnleyfootballclub.com/	76	1,911	60,884	4,336
CFC	Chelsea	https://www.chelseafc.com/	64	2,433	61,506	4,268
CPFC	Crystal Palace	http://www.cpfc.co.uk/	74	1,184	49,880	3,841
EFC	Everton	http://www.evertonfc.com/	76	2,623	70,641	5,364
HT	Huddersfield Town	https://www.htafc.com/	38	1,441	45,388	4,121
HC	Hull City	http://www.hullcitytigers.com/	37	923	24,470	2,132
LC	Leicester City	http://www.lcfc.com/	37	920	23,412	2,480
LFC	Liverpool	http://www.liverpoolfc.com/	62	1,763	41,089	3,446
MC	Manchester City	http://www.mancity.com/	76	2,204	52,690	4,251
MU	Manchester United	http://www.manutd.com/	76	1,772	51,000	4,204
MFC	Middlesbrough	http://www.mfc.co.uk/	38	1,018	25,161	2,747

(Continued)

14. FC in the abbreviations stands for “Football Club”, while AFC stands for “Association Football Club”, with the exception of Arsenal Football Club (AFC).

Table A. (Continued)

Corpus id. / Club abbr. ¹⁴	Team	URL	# of reports	# of s-units (AntConc counts)	# of tokens (AntConc counts)	# of types (AntConc counts)
NU	Newcastle United	https://www.nufc.co.uk/	38	813	20,265	2,409
SFC	Southampton	https://southamptonfc.com/	75	1,847	51,658	3,987
SC	Stoke City	http://www.stokecityfc.com/	73	1,164	40,102	3,713
SAFC	Sunderland	https://www.safc.com/	38	909	27,144	3,133
SCAFC	Swansea City	http://www.swanseacity.net/	76	1,853	41,754	3,283
TH	Tottenham Hotspur	http://www.tottenhamhotspur.com/	75	2,647	61,895	3,698
WFC	Watford	https://www.watfordfc.com/	74	2,338	54,508	4,463
WBA	West Bromwich Albion	http://www.wba.co.uk/	74	968	25,917	2,510
WHU	West Ham United	http://www.whufc.com/	70	1,360	36,743	3,202
Total					991,468	

Table B. Overview of *Eliteserie* teams, URLs and s-units in the individual corpus files¹⁵

Corpus id. / Club abbr.	Team	URL	# of reports	# of s-units (AntConc counts)	# of tokens (AntConc counts)	# of types (AntConc counts)
AaFK	Aalesund	http://www.aafk.no/	30	1,060	16,903	2,563
BG	Bodø-Glimt	https://www.glimt.no/	18	553	6,593	1,281
SKB	Brann	https://www.brann.no/	30	989	10,978	1,737
KBK	Kristiansund	http://www.kristiansundbk.no/	56	1,922	26,421	3,156
LSK	Lillestrøm	http://www.lsk.no/	59	1,836	21,795	2,631
MFK	Molde	https://www.moldefk.no/	24	519	6,951	1,469
OBK	Odd	http://www.odd.no/	58	1,957	23,975	3,721
RF	Ranheim	http://www.ranheimfotball.no/	30	1,107	14,386	2,008
RBK	Rosenborg	http://www.rbk.no/	58	1,050	16,723	2,138
SaF	Sandefjord	http://www.sandefjordfotball.no/	55	1,192	18,081	2,634
S08	Sarpsborg 08	http://www.sarpsborg08.no/	60	1,918	25,596	3,206
SoF	Sogndal	http://www.sogndalfotball.no/	29	384	4,900	1,185
STB	Stabæk	http://www.stabak.no	54	1,095	15,763	2,652
IKS	Start	http://www.ikstart.no/	30	1,435	16,273	1,954
SIF	Strømsgodset	http://www.godset.no/	57	1,448	22,940	2,512

(Continued)

15. One Norwegian club did not publish any match reports online (FK Haugesund) and is therefore not represented in the ENMaRC. FK in the abbreviations stands for *footballklubb* 'football club', BK stands for *ballklubb* 'ball club', IF and IL stand for *idrettsforening/idrettslag* 'sports club', and F stands for *fotball* 'football'.

Table B. (Continued)

Corpus id. / Club abbr.	Team	URL	# of reports	# of s-units (AntConc counts)	# of tokens (AntConc counts)	# of types (AntConc counts)
TIL	Tromsø	http://www.til.no/	57	2,016	26,504	2,489
VFK	Viking	http://www.viking-fk.no/	30	1,120	17,551	2,464
VIF	Vålerenga	http://www.vif-fotball.no/	57	1,393	22,065	2,898
Total					314,398	

Cross-disciplinary and cross-linguistic uses of tensed verb phrases in the methods sections of master's theses

Sylvi Rørvik

Inland Norway University of Applied Sciences

This study examines the distribution and functions of tensed verb phrases in the methods sections of master's theses produced at Norwegian universities. The material comprises texts written in both English and Norwegian within the disciplines of chemistry, linguistics, and sociology. The method involved manual identification of tensed verb phrases, as well as a more detailed investigation of their rhetorical functions. The findings indicate that there are cross-linguistic and cross-disciplinary differences in both distribution and function, in line with findings from previous studies of student and professional academic writing. These findings have implications for the teaching of academic writing, which should reflect the conventions of the disciplinary communities of which the student writers are aspiring members.

1. Introduction

Figures from the OECD and the EU show that the number of students attaining a higher-education degree has increased in the past 20 years, in a great number of countries around the world (OECD 2000, European Commission 2019, OECD 2019). Simultaneously, the proportion of international students, i.e. those studying abroad, has increased (OECD 2000, European Commission 2019, OECD 2019). There has, not surprisingly, been a corresponding upward development in the number of higher-education programs taught in English in countries where English is not an official language, as indicated by figures for various European countries (Maiworm & Wächter 2002; Wächter & Maiworm 2014). Given the higher degree of student mobility referenced above, it seems fair to assume that a similar development in the frequency of programs taught in English has taken place outside of Europe as well. We may thus tentatively conclude that a greater proportion of

students now write their master's theses in English, in a variety of subjects. This would hold true for Norway, as well, where the position of Norwegian as an academic language is being weakened, and the number of master's theses written in English is increasing (Gregersen 2014; Schwach & Elken 2018).

Given this development, there is a need for the explicit teaching of discipline-specific academic writing in both English and Norwegian in higher education in Norway, as well as in other countries, since research has shown that students do not adapt to academic writing conventions by osmosis alone (cf. e.g. Berry 1995), and since there is ample evidence for the existence of disciplinary differences in academic writing that make general academic writing courses less than ideal (cf. e.g. Hyland 2000; Fløttum et al. 2006; Kuteeva & McGrath 2013; Salas 2015; Bruce 2016; Jiang 2017). Furthermore, the teaching of academic writing in English must be suitable also for students who choose to write in English even though it is not their first language, and even though they do not have any formal qualifications in English beyond compulsory schooling. Students belonging to this group in Norway also have to be aware of the potential problems caused by cross-linguistic differences between English and Norwegian (see e.g. Fløttum et al. 2006; Nordrum 2007; and Dahl 2009). To date, relatively few studies have focused on academic writing in Norwegian institutions of higher education, and we therefore have comparatively little research on which to base the teaching of academic writing for these student groups. The present study is a step towards filling this gap, by investigating the language in master's theses written in English and Norwegian at Norwegian institutions of higher education.¹ A further aim is to contribute to the body of knowledge on academic writing, as well as contributing a basis for further exploration of non-expert academic writing produced in Norway.

The specific focus of this study is the use of tensed verb phrases (VPs) in the methods sections of master's theses.² Methods sections were selected as suitable material for an investigation of the use of tensed VPs because it could reasonably be assumed that they would contain procedural descriptions, which would necessitate a choice of tense to signal chronology. Previous research has shown that the choice of tense differs between different discourse segments in expert academic

1. The phrasing is deliberate, as the compilation process of the corpus that forms the basis for the present investigation did not include the collection of metadata on the authors' language background. The majority of students in higher education in Norway are not international students, so it may be assumed that they have Norwegian as their L1 and English as an L2, and consequently that the English material employed in this study was produced by L2 writers. See Section 3 for more details on the corpus compilation process.

2. The term 'use' is here employed as an umbrella term for distribution and function, see further Section 3.

writing, i.e. in published research articles (see de Waard & Maat (2012) for an overview), and there are indications that methods sections are one of the sections of academic texts where disciplinary differences and/or differences between “hard” and “soft” fields of science are most evident (Swales 1990: 175–176). Thus, in order to cover a range of disciplines across the hard/soft spectrum (see also Becher 1994), the material comprises methods sections from master's theses within chemistry, linguistics, and sociology. The research questions are as follows:

1. To what extent can disciplinary differences be identified in the distribution and functions of tensed VPs in the methods sections of master's theses written in English and Norwegian at Norwegian institutions of higher education?
2. To what extent do any such cross-disciplinary differences correlate with cross-linguistic differences in the distribution and functions of tensed VPs in the methods sections of master's theses written in English and Norwegian at Norwegian institutions of higher education?

The remainder of this chapter is structured as follows: Section 2 presents an overview of some relevant previous work, with a focus on disciplinary variation in student writing in Section 2.1 and on the distribution and function of tense in academic writing in Section 2.2. In Section 3, the material and method employed in the present investigation is presented, including details on the corpus compilation process, while Section 4 contains the results of the present investigation, as well as a discussion of them. The final section, Section 5, contains a summary and some concluding remarks.

2. Previous research

The present chapter aims to contribute to the body of knowledge regarding disciplinary variation in student academic writing, as manifested in the use of tensed VPs in methods sections. The overview of previous research presented in this section has therefore been structured with these two main areas in mind, i.e. studies focusing on disciplinary variation in student academic writing, and studies focusing on the distribution and functions of tensed verb phrases.

2.1 Disciplinary variation in student academic writing

The studies discussed in this section all focus on disciplinary variation in student academic writing in English. The overview is of course not exhaustive, but is intended to include evidence of disciplinary variation in a wide range of areas and among various student groups as this has emerged in studies carried out in the

past 20 years. Some studies are based entirely on native-speaker material, while others include texts written by both native and non-native speakers. The overview presented here is organized according to the linguistic features examined, such that studies with a similar linguistic focus have been grouped together. This resulted in four categories or topic areas: stance, citations/attribution, first-person pronouns, and lexical bundles/vocabulary.

The first topic area included here, stance, is represented by four studies. The first two of these deal with texts written by students whose first language is English, more specifically with master's theses in politics/international relations and doctoral dissertations in materials science (Charles 2003, 2007). Charles (2003) found that the corpus of texts within politics/international relations exhibited higher frequencies of nouns marked for stance than was found within materials science. Similarly, Charles (2007) examined the construction of stance through nouns followed by *that* and a complement clause, and found that this pattern was more frequent in politics/international relations than in materials science. Chan (2015), on the other hand, investigated acknowledgements in PhD dissertations written by non-native speakers of English, namely Hong Kong Chinese students. The focus was on lexico-grammatical devices that express stance, and the findings showed that stance features were used more frequently in the soft disciplines investigated (applied linguistics, business studies, public administration) than in the hard disciplines (biology, computer science, electrical engineering). Finally, Lancaster (2016) focused on expressions of stance in argumentative papers produced by (presumably) native-speaker students taking courses in either economics or political science. It was found that the disciplines favored different stance expressions: in economics there was more frequent use of hedges, disclaim markers, boosters, and self-mentions, while in political science there was more frequent use of attitude markers. We may thus conclude that there is ample evidence for disciplinary variation in the expression of stance, both in the frequency of such features and in the types of features employed in different disciplines.

We will now briefly look at three studies that may serve to illustrate disciplinary variation in the use of citations/attribution. Firstly, Charles (2006) used the same corpus as in the 2003 and 2007 studies discussed in the previous paragraph, i.e. master's theses in politics/international relations and doctoral dissertations in materials science, to investigate phraseological patterning in reporting clauses used to reference the research of others, and found some evidence of disciplinary differences in grammatical subject and verb choice. Samraj (2008) examined the use of citations in 24 master's thesis introductions produced at a US university. These theses were divided equally between the fields of philosophy, biology, and linguistics, and the results showed that citations were more frequent in biology than in linguistics, with the frequency in philosophy even lower than that found

in linguistics. As a final example of previous research on citation practices, we may include Ädel and Römer's 2012 study of disciplinary variation in the frequency and form of attribution in texts from MICUSP (*Michigan Corpus of Upper-level Student Papers*). Their study included texts from 10 disciplines, and showed that, in general, attribution was more frequent in the humanities, followed by the social sciences, biology, and health sciences, and least frequent in the physical sciences and engineering. Ädel and Römer's results seem to contradict those reported in Samraj's study, but it should be remembered that the text types examined differ in these two studies. It may well be that the expectations are different in master's theses than in the shorter texts examined by Ädel and Römer, or it may be that introductions have a higher frequency of attributions than is found in later sections of master's theses, so that Samraj's results might have been more similar to Ädel and Römer if the entire theses had been examined.

There have been numerous studies of the use of first-person pronouns in academic prose, so here I will just mention a few examples that illustrate disciplinary variation in the use of self-mention by student writers. The 2008 study by Samraj mentioned above also included an examination of the use of first-person pronouns in the introductions of master's theses, and found that these were most frequent in philosophy and least frequent in biology, with linguistics occupying the intermediate position. Lancaster (2016), referenced above in relation to stance) found that economics students used self-mentions more frequently than political-science students. Finally, Crossley et al. (2017), in a study lexical and cohesion differences among disciplines in MICUSP, found that the authors of engineering texts made more frequent use of pronouns than authors of texts in the sciences (the disciplines examined included biology, physics, mechanical engineering, and industrial and operations engineering). The results of these three studies seem to point in the same direction, namely that there is a continuum in terms of the frequency of first-person pronouns from the highest frequency in the soft disciplines to the lowest frequency in the hard sciences.

The final category that will be discussed is 'lexical bundles/vocabulary'. Cortes (2004) used both published research articles and student papers in history and biology to investigate students' use of target lexical bundles identified in the published texts. The findings showed that the experts used a wider range of bundles in biology than in history, but that the bundles performed similar functions in the two disciplines. In the student-produced texts, the frequency of target bundles was low, and the bundles that did occur did not always fulfil the functions that they did in the expert texts. Similarly, Hyland (2008) examined disciplinary variation in the frequency and use of recurrent four-word bundles in research articles, doctoral dissertations, and master's theses from four disciplines: electrical engineering, biology, business studies, and applied linguistics. The student texts were written by

L2 writers of English from five Hong Kong universities. Hyland found that there were disciplinary differences in the frequency and range of bundles employed, as well as in the frequency of the functions for which they were used. There was little overlap between fields in terms of which bundles occurred in the texts. Additional evidence of disciplinary differences in vocabulary-related features was identified in a study by Durrant (2013), who used student texts from 29 disciplines taken from the BAWE (*British Academic Written English*) corpus. The focus of the study was the extent to which students across disciplines used specialized or generic vocabulary, and whether there were clusters of students with similar vocabulary needs. The results showed that there were large disciplinary differences in high-frequency content words, with less than 50% being generic vocabulary. In addition, less than 50% of the generic uses were similar across student groups. A further study by Durrant (2015) examined disciplinary variation in student writing as manifested in the use of recurrent four-word sequences. It was found that in the BAWE texts included, which came from 24 disciplines, there were four main clusters: humanities and social science, science and technology, life sciences, and commerce. There was very little overlap between the first two groups, i.e. humanities and social science on the one hand and science and technology on the other hand, while the two final groups (life sciences and commerce) occupied an intermediate position with characteristics shared with both of the first two groups. As evidenced by these four studies on vocabulary-related topics, there are indications pointing to the existence of disciplinary differences in this area as well as in the three other areas discussed previously in this section.

2.2 The distribution and functions of tensed verb phrases

In order to facilitate cross-references to the findings from previous research in the discussion of the results from the present study, this presentation of selected earlier works on the distribution and function of tensed verb phrases has been divided into two subsections: Section 2.2.1 deals with previous findings on tense distribution, while Section 2.2.2 deals with the rhetorical functions of tensed verb phrases.

2.2.1 *Tense distribution and verb-phrase form*

Studies of tense distribution in academic writing seem, with some exceptions, to suggest that the past tense predominates in English academic prose. Hanania and Akhtar (1985), for instance, examined verb tenses in MSc theses in biology, chemistry, and physics. In the methods sections of these theses, they found that approximately 2/3 of verbs were in the past tense, and the remainder were in the present tense (1985: 52). The results for physics were slightly different from those

in biology and chemistry, however, in that this discipline contained a greater proportion of present-tense verb phrases than did the other two (ibid: 54). Similarly, Salager-Meyer (1992: 97) found that the past tense accounted for a little more than half of the verb phrases in medical abstracts, with the present tense accounting for approximately 1/3 of instances, and Li and Ge (2009) examined medical research articles and found that the past tense accounted for approximately 2/3 of tensed verb phrases, and the present tense for about 25% of instances (2009: 99). As regards academic writing produced by students, we may look to a comparison of argumentative essays written by native-speaker and non-native-speaker students at four US universities (Hinkel 2004), where it was shown that L2 learners from a range of L1 backgrounds had significantly higher frequencies of past-tense verbs than native-speaker students. However, Biber et al., in their overview of register distribution of tense showed that academic prose overall had a higher frequency of present-tense verbs than of past-tense verbs (1999: 456), and this might indicate that the other findings discussed here were influenced by the academic discipline from which the material was taken in each case. Support for this hypothesis can be found in a 2001 study by Taylor, who investigated tense use in 18 research articles from three humanities disciplines: English, history, and philosophy. The results of this study showed clear evidence of disciplinary variation even within the humanities, in that the simple present accounted for more than 70% of tensed verb phrases in English and philosophy, while in history the past tense accounted for a similar proportion.

Some other differences between the aforementioned studies emerge when we look at their findings regarding complex verb phrases. Taking their three investigated disciplines combined, Hanania and Akhtar (1985: 52) found that in the student theses 68.5% of tensed verb phrases were in the passive voice, with the vast majority of these being in the past tense. Again, however, physics was found to be different from biology and chemistry, in that the percentages for chemistry and biology were approximately 80% passives with the majority being past passives, while physics had approximately 40% passives, with the majority being present passives (ibid: 54). On the other hand, Salager-Meyer (1992: 97–98) found that both tenses were used predominantly in the active voice in medical abstracts, and also noted that complex verb phrases explicitly marked for aspect were very infrequent, with percentages below 10%. Both of these findings are in line with those of Biber et al. (1999: 476), who found that the vast majority of verb phrases were active, but in academic prose passive voice accounted for approximately 25% of verb phrases, and the majority of instances were short passives, where no agent is mentioned (ibid: 477). As regards aspect, the results reported in Biber et al. showed that the vast majority of verb phrases in academic prose were not explicitly marked for aspect, and that in those cases where the verb phrase was marked for

either perfect or progressive aspect, it was overwhelmingly the present perfect or the present progressive (ibid: 461–462). Taylor (2001) did not include voice in her study but found that all combinations of tense with explicitly marked aspect, i.e. both present and past perfect/progressive, were very rare, with percentages below 4% for all combinations. Hinkel's (2004) study did include voice as well as perfect aspect and showed that L2 writers had lower frequencies of complex VPs (i.e. passive voice/perfect aspect) than L1 writers. Further support for the low frequency of the present perfect can be found in the study by Li and Ge (2009: 99), where the present perfect accounted for approximately 6% of tensed verb phrases in medical research articles.

The studies discussed thus far point in the direction of both similarities and differences between student writing and expert academic prose in English. At both levels, the general tendency is for the past tense to be the most frequent tense choice (albeit with a potential for disciplinary bias influencing this finding), but student writers use the passive voice more frequently than established scholars do, although L2 student writers exhibit lower frequencies of complex VPs than L1 students. Finally, verb phrases explicitly marked for aspect are rare.

2.2.2 *Rhetorical functions of the present and past tense*

This overview of rhetorical functions identified in previous studies comprises some of those studies already discussed in Section 2.2.1, as well as two additional ones which examine the function of the tenses, but without including any (comparable) frequency data.

In the MSc theses examined by Hanania and Akhtar (1985), it was found that general discussions of mathematical models tended to be coded in the present tense and active voice, while descriptions of practical work tended to be coded in the past tense and passive voice (ibid: 54, 57). This is very similar to the findings presented by Malcolm, in a study of scientific articles within pediatrics: general statements occurred in the present tense, while references to specific studies occurred in the past tense (1987: 36). Malcolm further found that references to nonlinguistic information such as charts and illustrations, as well as summaries of the content of an article, tended to occur in the present tense (ibid: 35). The distinction between general truths, which tended to be expressed in the present tense, and specific procedures, which tended to be described in the past tense, was emphasized also by Salager-Meyer (1992: 106), and Biber et al stated that the present tense was used to express “information that is generally valid” (1999: 458).

Taylor (2004) included rhetorical functions in the study of present and past tense/aspect combinations. No frequencies were provided, but general trends were discussed. First, the simple present was typically used to state the purpose of an article and to express an argument, and to express the writer's comments, while

the simple past was typically used to recapitulate points previously mentioned. Second, two other functions could occur either in the simple present or in the simple past, depending on the type of reference: references to previous literature occurred in the simple present if the reference was to established knowledge, but in the simple past if the reference was to specific activities and events in the past; and references to primary sources were in the simple present if they were references to unchanging works of art or timeless opinions, but in the simple past if they were references to biographical details (ibid: 74).

A final interesting study of methods sections to be included here is Lim's 2006 investigation of methods sections in management research articles. This study focused on the rhetorical moves in methods sections, and how these were realized by means of lexical and syntactic choices. Tense distribution was not systematically commented on, but the study provided a useful addition to the list of rhetorical functions that one might expect to find in a methods section. Among the moves and steps proposed by Lim as features of methods sections in management research articles were "describing the data collection procedures", "describing the sample", "explaining the method", and "previewing results" (2006: 287). Lim did not provide frequency data for these features, but rather coded the articles he investigated for the presence or absence of these features. It will therefore not be possible to compare the results of the present investigation with those presented by Lim, but it will be of interest to note the extent to which these functions can be identified in the present material.

The previous studies discussed in this section have all contributed to the classification scheme employed in the present study, which will be introduced in Section 3.2.2.

3. Material and method

This section is divided into two subsections, the first of which describes the corpus compilation and sampling process, and the second of which outlines the analytical approach employed.

3.1 Corpus compilation and sampling

The data employed in the present investigation stems from a corpus intended to serve as the basis for an ongoing project entitled "Postgraduate Academic Writing in Norway" (PAWN). The main purpose of this project is to map lexico-grammatical features of student academic writing in Norway, in both Norwegian and English. The principal stages of data compilation took place in the period from May through August of 2018, and the procedure was as follows:

1. Master's theses and doctoral dissertations written in Norwegian and English in a range of disciplines³ that had been produced at Norwegian institutions of higher education between 2008 and 2018 were downloaded from the institutions' open-access digital repositories. In order to ensure that the maximum number of institutions possible was represented, and to avoid the potential bias caused by over-representation from individual institutions, the quota for each institution was limited to 25 texts per language per discipline, and these were collected starting from the most recent texts available, i.e. from 2018.
2. The texts were converted from pdf-format to txt-format, using the AntFile-Converter version 1.2.1 (Anthony 2017). Texts that were not successfully converted were replaced, if possible.⁴

The resulting corpus comprised 1,288 master's theses written in English, and 1,588 theses written in Norwegian, from a total of 13 institutions of higher education in Norway. From this pool, the texts included in the present study were sampled, and it is thus important to be aware of the limitations inherent in the corpus compilation process. Principally, these relate to the following factors: a) no metadata about the authors was available, beyond their names, and b) it follows from point a) that it is not possible to determine the authors' L1, as a person's name is not a sufficient indicator of their first language. Based on the fact that the majority of students in higher education in Norway are domestic students, we may assume that the majority of texts represented in the corpus were produced by authors whose first language is Norwegian, and this applies also to those texts that had been written in English. This should be borne in mind when interpreting the results but does not constitute a significant drawback for the present study, as the research questions do not presuppose native-speaker status among the authors of the texts investigated (cf. Section 1).

For the present study, texts within the disciplines chemistry, linguistics, and sociology were randomly sampled for explicitly labeled methods sections until a total of 25 was reached in each language in each discipline (150 in total, divided evenly between the languages).⁵ An overview of the material is presented in Table 1.

3. Eleven disciplines were included: biology, chemistry, comparative literature, economics, education, history, law, linguistics, medicine, psychology, and sociology. These disciplines were selected to ensure that both hard and soft fields were represented but were limited by the fact that the purpose of the project required disciplines where texts were being written in both Norwegian and English.

4. For funding reasons, the corpus is not yet publicly available.

5. To ensure the anonymity of the authors, it is not possible to include a list of the thesis titles here.

Table 1. Overview of the material: number of words in the methods sections of the selected corpora

	Corpus	Number of words	Total words
English subcorpora (75 texts)	Chemistry	69,295	287,752
	Linguistics	99,925	
	Sociology	118,532	
Norwegian subcorpora (75 texts)	Chemistry	60,618	271,225
	Linguistics	110,313	
	Sociology	100,294	
Total			558,977

The total extent of the material is in excess of 550,000 words, relatively evenly divided between the two languages. The disciplinary subcorpora differ in size, however, and this is probably due to disciplinary conventions, since one might expect more concise procedural descriptions in chemistry than in the two other disciplines. The comparability of these subcorpora results from the fact that they comprise the same text type, i.e. explicitly labeled methods sections, from the same genre, i.e. master's theses. Cross-linguistically, the comparison is not between L1 and L2 speakers of the same language, but between Norwegian produced by what are probably native speakers and English produced by what are probably non-native speakers with Norwegian as their L1 (cf. the discussion earlier in this section). The relevance of this comparison is that these texts are sampled from the relatively recent production of student academic writing in Norway, and hence the results of the present investigation may serve both as a starting point for pedagogical application in courses in academic writing, as well as a basis for comparison with both expert academic writing and L1 student academic writing in English.

3.2 Analytical approach

The research questions that formed the starting point for the present investigation were concerned with both the distribution of tensed verb phrases and their rhetorical function. In order to answer these questions, the study comprised two stages, and the description of the analytical approach has been organized according to those same two stages: stage 1, described in Section 3.2.1, comprised the identification of tensed verb phrases, while stage 2, described in Section 3.2.2, comprised a more detailed examination of the form and rhetorical functions of a subset of tensed verb phrases.

3.2.1 *Identification of tensed verb phrases*

In this first stage of the investigation, all tensed verb phrases in the material were identified manually. The tense systems of English and Norwegian are similar

in that verbs are inflected for present and past tense (Faarlund et al. 1997: 468; Biber et al. 1999: 453), but differ in that Norwegian modal auxiliaries, unlike English ones, have inflectional forms and can occur in both present and past tense (Faarlund et al. 1997: 491, 526; Biber et al. 1999: 453). In order to ensure that the cross-linguistic comparison carried out as part of the present study was not skewed by this typological difference, verb phrases containing modal auxiliaries were not included in the material. The manual identification of tense forms resulted in a total of approximately 45,000 tensed verb phrases, with a fairly even distribution between the two languages, as can be seen in Table 2.

Table 2. Overview of tensed verb phrases in the material

	Corpus	Tensed verb phrases	Total tensed verb phrases
English subcorpora (75 texts)	Chemistry	4,633	20,815
	Linguistics	7,103	
	Sociology	9,079	
Norwegian subcorpora (75 texts)	Chemistry	5,365	24,141
	Linguistics	9,669	
	Sociology	9,107	
Total			44,956

Once the tensed verb phrases had been identified, the disciplines and languages were compared, to determine their differences and similarities as regards the distribution of past- and present-tense verb phrases. This comparison was aided by statistical testing, which was conducted in RStudio (RStudio Team 2015) and consisted of one-way ANOVAs with Tukey post hoc tests.⁶

3.2.2 Identification of VP form and rhetorical function

In the second stage of the investigation, a subset of 3,000 tensed verb phrases were subjected to a more detailed analysis, which included verb-phrase complexity and rhetorical function. These verb phrases were randomly selected from each subcorpus, but the selection was limited to 250 verb phrases in the present tense and 250 verb phrases in the past tense from each of the six subcorpora. Although the total

6. By using ANOVA instead of multiple testing by means of tests comparing just two groups, one avoids the increased risk of Type I errors associated with repeated testing. The Tukey post hoc test is run in cases where the ANOVA shows that there is one or more significant differences, in order to show which groups significantly differ.

of 3,000 verb phrases included in this second stage of examination only accounted for approximately 7% of the total number of verb phrases, it was hypothesized that major trends in the material would still emerge from the analysis.

The study of verb-phrase complexity included the following features: simple vs. complex verb phrase (though with the exception of verb phrases containing modal auxiliaries, cf. Section 3.2.1), aspect, and voice, although for space reasons only aspect and voice are commented on in detail in the results section. The distinction between simple and complex verb phrases is relatively unproblematic as concerns a cross-linguistic comparison of English and Norwegian when verb phrases containing modal auxiliaries have been excluded: a simple verb phrase consists of only one word, while a complex verb phrase consists of more than one word (Quirk et al. 1985: 151; Biber et al. 1999: 453; Faarlund et al. 1997: 520). However, there are some cross-linguistic differences with regard to complex verb phrases, specifically to do with the explicit marking of aspect and passive voice. The present study includes, for English, the perfect and progressive aspects, which are marked by a combination of the auxiliary verb 'have' and the past participle of the lexical verb (perfect) and by a combination of the auxiliary 'be' and the present participle of the lexical verb (progressive) (Quirk et al. 1985: 151; Biber et al. 1999: 406). Norwegian does not have the progressive aspect, so only the perfect is relevant, and this is marked by a combination of the auxiliary *ha* ('have') with the past participle of the lexical verb (Faarlund et al. 1997: 520). Verb phrases that are marked for the passive in English usually consist of the auxiliary 'be' and the past participle of the lexical verb (Quirk et al. 1985: 151, Biber et al. 1999: 475), while in Norwegian there are two types: the most frequent one, which results in a complex verb phrase, with a combination of the auxiliary verbs *bli* ('become') or *være* ('be') combined with the past participle of the lexical verb, and one that results in a simple verb phrase, where the passive is marked by adding an 's' ending to the lexical verb (Faarlund et al. 1997: 513, 523).

The categories employed in the analysis of the rhetorical function of the verb phrases were inspired by several of the previous studies discussed in Section 2.2.2 (especially Malcolm 1987 and Lim 2006), and include the following nine categories, representing features typically occurring in a methods section: 'table/diagram', 'focus of chapter', 'summary of discussion', 'previous research', 'sample description', 'data collection', 'preview of results', 'method employed', and 'general method description'. Each of these will now be illustrated by means of an example from the material.⁷

7. Examples are reproduced exactly as they occur in the corpora, including any errors, and are followed by a label referencing the subcorpus from which the example originates. Relevant verb phrases have been underlined.

The category called ‘table/diagram’ refers to clauses whose main function is metadiscoursal in that they point to a table, a diagram or another feature of the thesis as a written object. An example is the sentence in (1), where the author uses the present tense to describe the contents of a table:

- (1) Table 8 has an overview of the number of participants per language and language group. (English linguistics)

The second category, called ‘focus of chapter’, is similar to the first in that it points to a part of the thesis, but while the ‘table/diagram’ category deals with elements that have been added to the text, the ‘focus of chapter’ category is only used for references to the chapters of the thesis. Example (2) illustrates this usage:

- (2) Det første analysekapittelet (kapittel 4) tar utgangspunkt i musikkens funksjon i ungdommers hverdag. [‘The first analysis chapter (Chapter 4) takes as its starting point...’] (Norwegian sociology)

‘Summary of discussion’ has been used as the label for cases like that in example (3), where the writer sums up the content of the current chapter or section, before moving on to the next:

- (3) This chapter presented a brief description regarding the speakers in a general way. (English linguistics)

When the writer refers to studies conducted by other researchers, this has been coded as ‘previous research’. Example (4) illustrates this function:

- (4) En lignende studie ble utført som en del av TAUS-undersøkelsen på 1970-tallet i Oslo (Sandøy 1996: 97). [‘A similar study was carried out as a part of...’] (Norwegian linguistics)

It is not uncommon for writers to describe the sample that forms the dataset for the investigation, and this function, exemplified by the sentence in (5), is termed ‘sample description’:

- (5) The salts used was NaCl (Sigma Aldrich, 99.8% purity) and CaCl₂·2H₂O (Sigma Aldrich, 99% purity). (English chemistry)

In the present analytical framework, description of the samples has been considered a separate category from the one labeled ‘data collection’, which deals with all aspects of the procedure employed to obtain the data for the investigation. An example of this is given in (6):

- (6) Strandkrabbene ($n = 227$) ble samlet inn fra Kvitsøy, Sotra og Hardangerfjorden (Måge) i sør, samt Vesterålen (Guvåg) og Fleinvær i nord (figur 3.1.1). [‘The beach crabs ($n = 227$) were collected from Kvitsøy, Sotra and Hardangerfjorden (Måge) in the south ...’] (Norwegian chemistry)

Sometimes the methods sections also include a preview of the results of the investigation, as illustrated in (7):

- (7) Pulse lengths acquired were in the interval 6.7–7.5µs. (English chemistry)

The category 'method employed' comprises descriptions of all features of the research process that are not included in data collection, i.e. all methods employed at all stages of the analysis process, as well as methodological considerations and justification. Example (8) illustrates this category:

- (8) Accordingly I have selectively focused on those elements in the documents that I judged to be relevant to coordination.
(English sociology)

The final category, labeled 'general method description', was not on the original list of categories adapted from previous studies, but emerged from the data itself, based on cases like that in example (9):

- (9) En styrke ved kvalitative intervjuer (og case studier) mer generelt, er at man får samlet inn mer detaljert, fullstendig og troverdig informasjon, noe som kan være vanskeligere ved for eksempel spørreundersøkelser. ['A strength of qualitative interviews (and case studies) more generally is that one is able to gather more detailed, complete and credible information, which may be more difficult in the case of for instance questionnaires.']
(Norwegian sociology)

In this example, the author is not concerned with a description of the procedure employed in her own investigation, but rather presents a general reflection on the advantage of using qualitative interviews.

It should be noted, before we move on to the results, that no statistical calculations were carried out as part of this more detailed analysis, since it comprised only a small subset of the entire dataset. The results of this stage of the investigation should therefore be considered a first exploration of these features, to see if any of them warrant further investigation on a larger dataset.

4. Results and discussion

The structure of this section follows that of the methods section, such that the first subsection presents the results as regards tense distribution, while the second subsection presents the results of the more detailed analysis of a subset of the material.

4.1 Tense distribution

This section deals with the results relating to tense distribution in the material, both aggregate proportions for each subcorpus, and corpus-internal distribution.

Table 3 contains the total number of past- and present-tense verb phrases in each subcorpus.

Table 3. Distribution of past- and present-tense verb phrases in each subcorpus (raw figures and percentages of total verb phrases)

Corpus	Past tense		Present tense		Total number of tensed verb phrases
	Raw figures	%	Raw figures	%	
Eng chem.	2,612	56.4	2,021	43.6	4,633
Nor chem.	3,306	61.6	2,059	38.4	5,365
Eng ling.	2,914	41	4,189	59	7,103
Nor ling.	2,947	30.5	6,722	69.5	9,669
Eng soc.	3,812	42	5,267	58	9,079
Nor soc.	4,096	45	5,011	55	9,107
Total	19,687	43.8	25,269	56.2	44,956

Table 3 shows the following disciplinary characteristics within each language: Among the English subcorpora, we find the highest proportion of past-tense verb phrases in chemistry, while linguistics and sociology are nearly identical. The opposite is then true for present-tense verb phrases: nearly identical proportions in linguistics and sociology, which are higher than the one found in chemistry. Among the Norwegian subcorpora, we also find that past-tense verb phrases are most common in chemistry. They are second-most frequent in sociology, and least frequent in linguistics. Conversely, present tense is most common in linguistics, followed by sociology and chemistry. If we sum up the cross-linguistic differences and similarities, we have seen that past tense is most frequent in chemistry in both languages, accounting for more than 50% of tensed verb phrases, while for the other disciplines it is present tense that predominates. We also note that there are clearer disciplinary differences in the Norwegian material than in the English material. A comparison with the results from the previous studies discussed in Section 2.2.1 shows that the findings in the literature are probably discipline-specific to a large degree, since the predominance of the past tense found in medical research articles and medical abstracts (Salager-Meyer 1992 and Li & Ge 2009) matches those found for chemistry in the present material, but not those for linguistics and sociology. We may note also the similarity of the present results for chemistry to those presented by Hanania and Akhtar (1985). Biber et al's finding that the present tense predominates in academic prose in English (1999: 456) is borne out by the present material, although this might not have been the case if more disciplines from the hard sciences had been included in the present investigation, since

chemistry has a higher frequency of past-tense verb phrases in line with the other studies based on such material discussed above. On the other hand, Taylor (2001) found variability in tense distribution in within the humanities, so there may not be a clear-cut division between the hard and soft sciences regarding this feature.

The aggregate proportions for each language discussed thus far are based on figures for the subcorpora as wholes, and may therefore be misleading in the seemingly coherent picture they present of disciplinary differences and similarities. We will therefore now turn to a closer look at the corpus-internal distribution, to see if any new nuances emerge that can further inform our impression of tense use in the three disciplines and two languages.

Figure 1 shows the corpus-internal distribution of past-tense verb phrases in both languages (the distribution of present-tense verb phrases is a mirror image of Figure 1, and is therefore not included here).

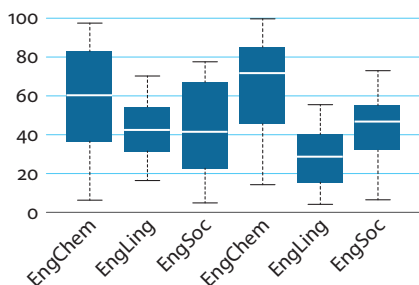


Figure 1. Proportion (in percent) of past-tense verb phrases in each subcorpus

What is immediately clear from Figure 1 is that there is a lot of corpus-internal variation in each of the six subcorpora. We should therefore interpret the aggregate results with some caution, and treat them as indicators of areas that need to be further explored in more extensive datasets. The tendency that can be extracted from these figures, however, is that the impression formed on the basis of the aggregate proportions is largely correct: in both languages, chemistry has the highest proportion of past tense and the lowest proportion of present tense, and the disciplinary differences are more prominent in Norwegian than in English, as is evident from the lesser degree of overlap between the boxes in the plots for the Norwegian subcorpora as compared to the English subcorpora. This conclusion is also supported by the statistical analysis, which indicates that the only significant difference between the three English disciplines is found between chemistry and linguistics ($p = 0.0293101$), whereas for Norwegian we find significant differences between all three subcorpora (chemistry vs. sociology: $p = 0.0007396$; chemistry vs. linguistics: $p = 0.0000000$; and sociology vs. linguistics: $p = 0.0305000$). This

may perhaps be linked to greater writing proficiency in Norwegian than in English, since Norwegian is probably the L1 of most of the authors (cf. Section 3.1): it is possible that the authors have a better grasp of disciplinary conventions in their first language than in their second language.

If we consider the three disciplines in a cross-linguistic perspective, we find that there are no significant differences between the English and Norwegian material within the disciplines of chemistry ($p = 0.4468671$) or sociology ($p = 0.930165$). However, there is a cross-linguistic difference between English and Norwegian linguistics ($p = 0.0043905$). Figure 2 shows the plot for past tense within linguistics in the two languages (as was the case with Figure 1, the plot for present tense is the mirror image of the plot for past tense). As can be seen from the plot in Figure 2, past tense is more frequent in English than in Norwegian, and the opposite is then true of present tense.

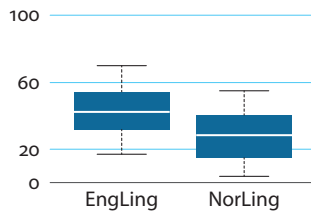


Figure 2. Cross-linguistic comparison of past-tense distribution within linguistics

The results thus far indicate that, of the three disciplines under investigation, chemistry is “the most different” in that it has a higher proportion of past-tense verb phrases, and a lower proportion of present-tense verb phrases than the other two disciplines. As regards linguistics and sociology, the picture is less clear, however: these disciplines are fairly similar in the English material, but different in the Norwegian material, since there are more present-tense verb phrases in linguistics and more past-tense verb phrases in sociology. Cross-linguistically, the only significant difference is between the two subcorpora containing linguistics texts. It is not possible on the basis of the tense-distribution numbers to find explanations for these differences, so we now turn to a more detailed look at a subset of the verb phrases in the material.

4.2 Verb-phrase form and rhetorical function

Section 4.2.1 presents a closer look at verb-phrase form in a subset of the material (cf. Section 3.2.2), while the rhetorical function of clauses is examined in Section 4.2.2.

4.2.1 *Verb-phrase form*⁸

The previous studies discussed in Section 2.2.1 above suggested a difference between student writing and expert writing in English, in the sense that student writers tended to use more complex verb phrases in the form of passive constructions (Hanania & Akhtar 1985) than were found in expert writing (Salager-Meyer 1992; Biber et al 1999; Li & Ge 2009). Perfect and progressive aspects were found to be rare in all text types and disciplines, however.

Table 4 shows the proportion (in percent) of verb phrases in each subcorpus containing either perfect or progressive aspect, or a combination of both aspects.⁹

Table 4. Proportion of verb phrases (in percent of the total in each tense category) marked for aspect(s) in each subcorpus

Corpus	Perfect		Progressive		Perfect progressive	
	Past	Present	Past	Present	Past	Present
Eng chem.	0	0.4	0.4	0	0	0
Nor chem.	0	4.4	N/A	N/A	N/A	N/A
Eng ling.	0	4	0	0	0	0
Nor ling.	5.2	15.2	N/A	N/A	N/A	N/A
Eng soc.	2.8	13.6	2.8	0.8	0	0.8
Nor soc.	0.8	20	N/A	N/A	N/A	N/A

In both languages and all three disciplines, the vast majority of verb phrases contain no aspectual marking. However, we find a slightly higher proportion of the present perfect in Norwegian linguistics and English and Norwegian sociology than in the other subcorpora, and these percentages account for nearly all the aspect marking in the material. The results of the present investigation are thus in line with those of previous studies, which also concluded that aspect marking is infrequent in academic prose (Biber et al. 1999; Li & Ge 2009).

The distribution of active and passive voice is shown in Table 5, where the percentages within each tense category add up to 100.

8. In all tables in this section, the percentages refer to the proportion of the 250 VPs investigated for each tense category in each subcorpus.

9. As Norwegian does not have the progressive aspect, there can be no cross-linguistic comparison with regard to this feature.

Table 5. Proportion of verb phrases (in percent of the total in each tense category) in each subcorpus in active vs. passive voice, with long passives presented separately

Corpus	Active		Long passive		Short passive	
	Past	Present	Past	Present	Past	Present
Eng chem.	14	49.2	2.8	0.8	83.2	50
Nor chem.	12	47.6	0	2.4	88	50
Eng ling.	52.4	75.6	2.8	0.8	44.8	23.6
Nor ling.	88.4	90	1.3	0.4	18	9.6
Eng soc.	85.6	87.6	0.8	1.6	13.6	10.8
Nor soc.	82	85.6	0.8	1.2	17.2	13.2

In English chemistry, the majority of past-tense verb phrases are in the passive, while for present tense there is a very nearly even distribution between active and passive. In English linguistics, there is a fairly even distribution in the past tense, while the majority of present-tense verb phrases are active. In English sociology, the majority of verb phrases are active in both tenses. Overall, a very small proportion of verb phrases occur in long passive constructions, i.e. ones with an expressed agent. The results for Norwegian chemistry are very similar to those for English chemistry, with the vast majority of past-tense verb phrases being in the passive voice, but with a nearly even distribution of passive and active voice in present-tense verb phrases. Norwegian linguistics is different from English linguistics, in that the majority of verb phrases are active in both tenses, and sociology in Norwegian, like in English, follows this pattern as well. Thus, it is English linguistics that stands out as different from the other subcorpora, in that it does not pattern with either the softer disciplines of sociology in both languages and Norwegian linguistics, nor with the harder discipline of chemistry. All disciplines in both languages are similar, however, in that the vast majority of passives are short passives, i.e. those where no agent is expressed.

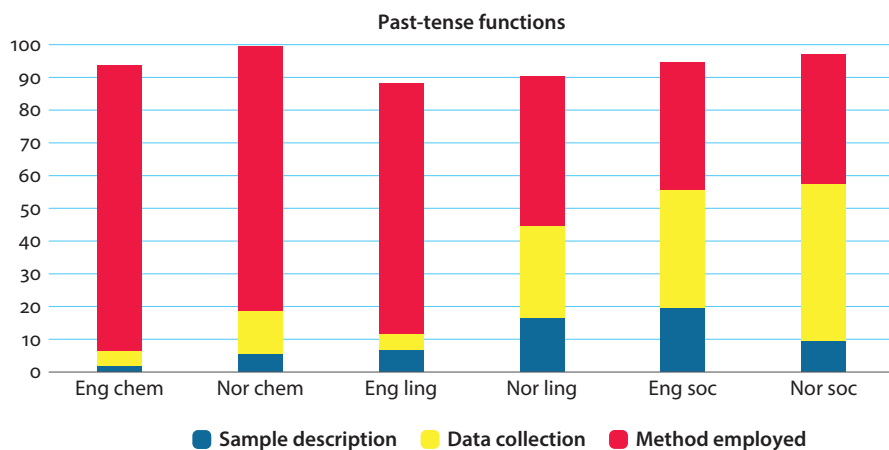
4.2.2 *Rhetorical function of clauses*

We now turn to the rhetorical function of clauses, starting with the results for clauses with past-tense verb phrases. The overall percentages for each function in past-tense verb phrases are presented in Table 6.

As can be seen from Table 6, there are only a few functions that occur with any noticeable frequency, i.e. in more than 10% of the clauses in at least one subcorpus. These are ‘sample description’, ‘data collection’, and ‘method employed’. Figure 3 presents a visualization of the proportions for these three categories.

Table 6. Rhetorical function (in percent) of clauses with past-tense verb phrases

	Eng chem.	Nor chem.	Eng ling.	Nor ling.	Eng soc.	Nor soc.
Table/diagram	0	0	0	0	0	0
Focus of chapter	0.4	0	0	0	0	0
Summary of discussion	0	0	5.6	0.4	0.4	0
Previous research	1.2	0	2.8	5.2	3.2	0.4
Sample description	2	5.6	6.8	16.4	19.6	9.6
Data collection	4.4	13.2	4.8	28.4	36	48
Preview of results	4.4	0.4	2	2.4	0.8	2
Method employed	87.2	80.8	76.4	45.6	38.8	39.6
General method description	0	0	0	0	0	0
Other	0.4	0	1.6	1.6	1.2	0.4
Total	100	100	100	100	100	100

**Figure 3.** Distribution (in percent) of frequent rhetorical functions of clauses with past-tense verb phrases

What is immediately clear from Figure 3 is that the majority of clauses with a past-tense verb phrase describe the method employed in the study. However, there are some disciplinary and cross-linguistic differences: in the English dataset, the function of 'sample description' is most frequent in sociology, and least frequent in linguistics and chemistry. Clauses describing the data-collection process are also much more frequent in sociology than in linguistics and chemistry, with an even

greater difference between sociology and the two other disciplines than is the case for 'sample description'. Clauses describing the method employed are most frequent in chemistry, less frequent in linguistics, and least frequent in sociology. In the Norwegian dataset, we find that 'sample description' accounts for a higher proportion of clauses in linguistics than in sociology, and only 2% of clauses in chemistry have this function. For 'data collection' we find by far the highest proportion in sociology, followed by linguistics. Again, the frequency in chemistry is very low. However, in chemistry the vast majority of clauses with past-tense verb phrases are used to describe the method employed, and the proportion for this function is nearly twice as high in chemistry as it is in linguistics, and more than twice as high as the proportion in sociology. Cross-linguistically, we find once again that chemistry texts behave similarly in the two languages, although the Norwegian chemistry texts have more description of data collection in the past tense than the English texts do. Within linguistics, there is a more even distribution between the three most frequent functions in Norwegian than in English, where 'method employed' accounts for a larger proportion of the clauses, and within sociology 'sample description' is more prevalent in English, 'data collection' is more prevalent in Norwegian, and 'general method description' occurs in nearly identical proportions in the two languages.

Table 7 shows that clauses with present-tense verb phrases occur with a greater range of functions than do clauses with past-tense verb phrases, and that the function of 'data collection', which was one of the three most frequent for clauses with past-tense verb phrases in Table 8, is extremely infrequent when a clause has a verb phrase in the present tense.

Table 7. Rhetorical function (in percent) of clauses with present-tense verb phrases

	Eng chem.	Nor chem.	Eng ling.	Nor ling.	Eng soc.	Nor soc.
Table/diagram	8.8	12.4	10	1.2	0.8	0
Focus of chapter	1.6	0.8	8.4	5.2	1.2	2.8
Summary of discussion	0	0	1.2	0	0.4	0.4
Previous research	0	0.8	4.4	7.2	15.6	7.2
Sample description	2	4.4	5.2	13.2	27.2	8.4
Data collection	0	2	0.4	0	0	4
Preview of results	0.8	0.4	8.8	0.8	0	0
Method employed	34.8	42	55.2	38.4	23.2	39.6
General method description	52	37.2	4.4	31.2	29.2	34.8
Other	0	0	2	3.8	2.4	2.8
Total	100	100	100	100	100	100

Five functions occur with a frequency of at least 10% of clauses in a minimum of one subcorpus, namely 'table/diagram', 'previous research', 'sample description', 'method employed', and 'general method description'. A visual representation of the frequencies for these categories is provided in Figure 4.

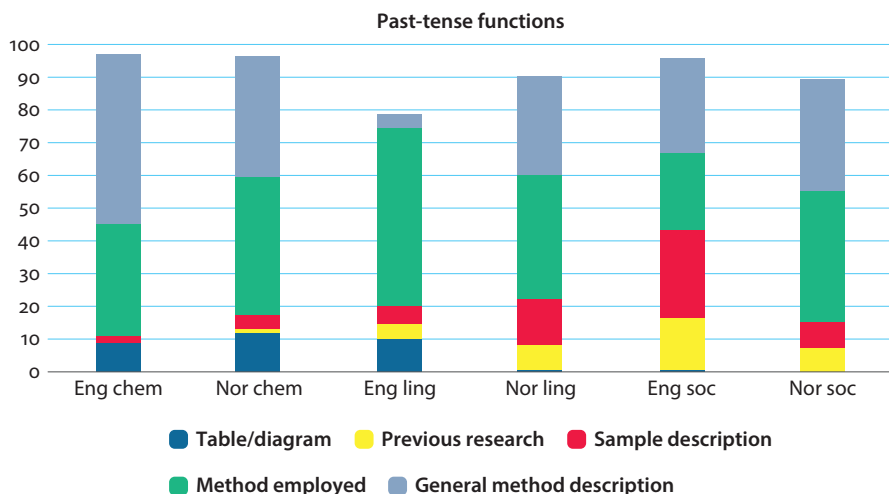


Figure 4. Distribution (in percent) of frequent rhetorical functions of clauses with present-tense verb phrases

If we compare the diagram in Figure 4 with that presented in Figure 3, perhaps the most striking difference is reduction in the frequency of 'method employed', which was more frequent with past-tense verb phrases than it is with present-tense verb phrases, which is in line with the findings from previous research (Hanania & Akhtar 1985; Salager-Meyer 1992). To a large extent, the gap left by the lower frequency of this function is filled by the function labeled 'general method description', but we also find references to tables and diagrams more frequently with present-tense verb phrases than with ones in the past tense (cf. Malcolm 1987). However, just as we saw with the functions of clauses with past-tense verb phrases, there are both disciplinary and cross-linguistic differences. In the English material, we find references to tables and diagrams in approximately 10% of clauses in chemistry and linguistics, but only 0.8% in sociology. References to previous research, on the other hand, are more frequent in sociology than in linguistics, and non-existent in chemistry. The function of 'sample description' is very infrequent in both chemistry and linguistics, but more frequent in sociology. Descriptions of the method employed account for more than half the cases in linguistics, but around 1/3 of clauses in chemistry and approximately 1/5 in sociology. Conversely, 'general method description' is very infrequent in linguistics, but accounts for approximately 1/3 of clauses in sociology

and a little over half of the clauses in chemistry. In the Norwegian material, references to tables and diagrams only occur with any frequency in chemistry. References to previous research, however, do not occur in chemistry, but are equally frequent at 7.2% in linguistics and sociology. Sample description is most frequent in linguistics, less frequent in sociology, and even less frequent in chemistry. All three disciplines in the Norwegian dataset are similar in their proportions of 'method employed' at approximately 40%. Similarly, when it comes to 'general method description', the proportions are fairly similar for all three disciplines, ranging from 31.2% in linguistics to 37.2% in chemistry. Cross-linguistically, we find that the two major functions of clauses with present-tense verb phrases, i.e. 'method employed' and 'general method description', are used more similarly across disciplines in Norwegian than they are in English. We also find that the two languages are similar in the sense that references to tables and diagrams are used in chemistry, and that references to previous research do not occur in the chemistry subcorpus. For the remaining features and disciplines discussed here, there are no clear patterns, which is perhaps an indication that a larger dataset is needed if one wishes to generalize about cross-linguistic differences and similarities. Alternatively, it might be the case that the use of tense in methods sections is not standardized to a great degree, but again this hypothesis needs to be tested on a larger dataset.

5. Concluding remarks

The main findings of this investigation can be summarized as follows:

- The present tense is more frequent than the past tense in linguistics and sociology, while the opposite is true for chemistry. This is the case in both languages, and the existence of disciplinary differences is not unexpected, given the findings reported by Hanania and Akhtar (1985) and Taylor (2001) (cf. Section 2.2.1).
- Short passives are more frequent in the past tense than in the present tense in all disciplines and in both languages, and this matches the results regarding past passives in previous studies (cf. Hanania & Akhtar (1985) concerning chemistry and biology). However, the majority of verb phrases in linguistics and sociology are active in both tenses and in both languages, while in chemistry the majority are passive in both tenses and in both languages. The preponderance of active verb phrases in linguistics and sociology is in line with Salager-Meyer's (1992) results based on medical abstracts.
- Explicit aspectual marking is rare, and this is also in line with previous studies (Salager-Meyer 1992; Biber et al. 1999; Taylor 2001; Li & Ge 2009).

- The range of rhetorical functions is greater in the present tense than in the past tense. The vast majority of clauses in the past tense describe the method employed, especially so in chemistry in both languages as well as English linguistics. In Norwegian linguistics and sociology in both languages, descriptions of the data collection process also account for a sizeable proportion of instances. In the present tense, the most frequent functions are general method description and descriptions of the method employed. This general tendency to have specific procedures recounted in the past tense, while more general statements are expressed in the present tense, reflects tendencies identified in previous research as well, for instance in the studies by Hanania and Akhtar (1985) and Salager-Meyer (1992; see further Section 2.2.2).

Caution should be employed in the interpretation of the results of the relatively small-scale detailed analysis conducted in the present study, at least until they have been confirmed by larger-scale studies. The results do, however, seem to indicate that there are both disciplinary and cross-linguistic differences, albeit with a degree of variation within each discipline (cf. Figure 1). Based on the present dataset as well as on previous studies (cf. Section 2), it would seem impossible to suggest a single set of “rules” that could apply to academic prose in general concerning to the choice of tense in methods sections, which points both to the necessity of discipline-specific writing courses, and to the necessity of such courses involving awareness training, during which the students are made aware of the range of rhetorical options open to them, and the effects of choosing each one. This is in line with Lim’s conclusion that students should “seriously consider the ways in which the communicative intentions are achieved linguistically” (2006: 303), and would go some way towards remedying the situation pointed out by Lillis (1997: 186), namely that “students know that there are certain rules governing what you can write in academic texts, but do not necessarily know what they are”.

In order to make such awareness training efficient, however, it is obvious that further research is necessary, to expand the knowledge on which it could be based. Taking the present investigation as the starting point, there are two obvious avenues for further research that could further increase our knowledge about the use of tensed verb phrases in academic writing, both of which could also be expanded through looking at other features than the ones included in the present study, e.g. tense variation relative to the lexical heads of verb phrases. The first concerns an expansion of the present study to also include the other sections of master’s theses. This would be helpful in identifying possible other patterns of use that students would also need to master. The second concerns a very necessary control mechanism in the form of expert academic writing, as one cannot take it for granted that

L1/L2 student writers necessarily choose the most felicitous ways of expressing themselves. Although research articles are not produced for the same reasons as master's theses, and may have other requirements governing them, they would nonetheless serve as a good yardstick for comparison if one accepts the premise that students are aspiring members of a disciplinary community.

References

- Ädel, Annelie & Römer, Ute. 2012. Research on advanced student writing across disciplines and levels. Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics* 17(1): 3–34. <https://doi.org/10.1075/ijcl.17.1.01ade>
- Anthony, Laurence. 2017. *AntFileConverter (Version 1.2.1)*. Tokyo: Waseda University. <<https://www.laurenceanthony.net/software>> (1 April 2021).
- Becher, Tony. 1994. The significance of disciplinary differences. *Studies in Higher Education* 19(2): 151–161. <https://doi.org/10.1080/03075079412331382007>
- Berry, Margaret. 1995. Thematic options and success in writing. In *Thematic Development in English Texts*, Mohsen Ghadessy (ed.), 55–84. London: Pinter.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
- Bruce, Ian. 2016. Constructing critical stance in University essays in English literature and sociology. *English for Specific Purposes* 42: 13–25. <https://doi.org/10.1016/j.esp.2015.10.005>
- Chan, Thomas Hon-Tung. 2015. A corpus-based study of the expression of stance in dissertation acknowledgements. *Journal of English for Academic Purposes* 20: 176–191. <https://doi.org/10.1016/j.jeap.2015.09.005>
- Charles, Maggie. 2003. 'This mystery...': a corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes* 2: 313–326. [https://doi.org/10.1016/S1475-1585\(03\)00048-1](https://doi.org/10.1016/S1475-1585(03)00048-1)
- Charles, Maggie. 2006. Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes* 25: 310–331. <https://doi.org/10.1016/j.esp.2005.05.003>
- Charles, Maggie. 2007. Argument or evidence? Disciplinary variation in the use of the Noun that pattern in stance construction. *English for Specific Purposes* 26: 203–218. <https://doi.org/10.1016/j.esp.2006.08.004>
- Cortes, Viviana. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23: 397–423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Crossley, Scott. A., Russell, David R., Kyle, Kristopher & Römer, Ute. 2017. Applying natural language processing tools to a student academic writing corpus: How large are disciplinary differences across science and engineering fields? *Journal of Writing Analytics* 1: 48–81. <https://doi.org/10.37514/JWA-J.2017.1.1.04>
- Dahl, Trine. 2009. Author identity in economics and linguistics abstracts. In *Cross-Linguistic and Cross-Cultural Perspectives on Academic Discourse* [Pragmatics & Beyond New Series 193], Eija Suomela-Salmi & Fred Dervin (eds), 123–134. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.193.08dah>

- Durrant, Philip. 2013. Discipline and level specificity in university students' written vocabulary. *Applied Linguistics* 35(3): 328–356. <https://doi.org/10.1093/applin/amto16>
- Durrant, Philip. 2015. Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics* 38(2): 165–193. <https://doi.org/10.1093/applin/amv011>
- European Commission. 2012. *Education and Training Monitor 2012*. Luxembourg: Publications Office of the European Union.
- European Commission. 2019. *Education and Training Monitor 2019*. Luxembourg: Publications Office of the European Union.
- Faarlund, Jan Terje, Lie, Svein & Vannebo, Kjell Ivar. 1997. *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.
- Fløttum, Kjersti, Dahl, Trine & Kinn, Torodd. 2006. *Academic Voices. Across Languages and Disciplines* [Pragmatics & Beyond New Series 148]. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.148>
- Gregersen, Frans. 2014. *Hvor parallellt. Om parallellspråkighet på Nordens universitet* [TemaNord 2014: 535]. Copenhagen: Nordisk Ministerråd.
- Hanania, Edith A. S., & Akhtar, Karima. 1985. Verb form and rhetorical function in science writing: A study of MS theses in biology, chemistry, and physics. *The ESP Journal* 4: 49–58. [https://doi.org/10.1016/0272-2380\(85\)90006-X](https://doi.org/10.1016/0272-2380(85)90006-X)
- Hinkel, Eli. 2004. Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research* 8(1): 5–29. <https://doi.org/10.1191/1362168804lr1320a>
- Hyland, Ken. 2000. *Disciplinary Discourses: Social Interaction in Academic Writing*. Harlow: Pearson.
- Hyland, Ken. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27: 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Jiang, Feng (Kevin). 2017. Stance and voice in academic writing: The “noun + that” construction and disciplinary variation. *International Journal of Corpus Linguistics* 22(1): 85–106. <https://doi.org/10.1075/ijcl.22.1.04jia>
- Kuteeva, Maria & McGrath, Lisa. 2013/2015. The theoretical research article as a reflection of disciplinary practices: The case of pure mathematics. *Applied Linguistics* 36(2): 215–235. <https://doi.org/10.1093/applin/amto42>
- Lancaster, Zak. 2016. Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes* 23: 16–30. <https://doi.org/10.1016/j.jeap.2016.05.006>
- Li, Li-Juan, & Ge, Guang-Chun. 2009. Genre analysis: Structural and linguistic evolution of the English-medium medical research article (1985–2004). *English for Specific Purposes* 28: 93–194. <https://doi.org/10.1016/j.esp.2008.12.004>
- Liddicoat, Anthony J. 2004. Grammar as a feature of text construction. Time and rhetorical function in French journal articles in biology. *Written Communication* 21(4): 316–343. <https://doi.org/10.1177/0741088304268806>
- Lillis, Theresa. 1997. New voices in academia? The regulative nature of academic writing conventions. *Language and Education* 11(3): 182–199. <https://doi.org/10.1080/09500789708666727>
- Lim, Jason Miin Hwa. 2006. Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes* 25: 282–309. <https://doi.org/10.1016/j.esp.2005.07.001>

- Maiworm, Friedhelm & Wächter, Bernd. 2002. *English-Language-Taught Degree Programmes in European Higher Education: Trends and Success Factors*. Bonn: Lemmens.
- Malcolm, Lois. 1987. What rules govern tense usage in scientific articles? *English for Specific Purposes* 6(1): 31–43. [https://doi.org/10.1016/0889-4906\(87\)90073-1](https://doi.org/10.1016/0889-4906(87)90073-1)
- Nordrum, Lene. 2007. English Lexical Nominalizations in a Norwegian-Swedish Contrastive Perspective. PhD dissertation, University of Gothenburg.
- OECD. 2000. *Education at a Glance 2000: OECD Indicators*. Paris: OECD Publishing. <https://doi.org/10.1787/eag-2000-en>
- OECD. 2019. *Education at a Glance 2019: OECD Indicators*. Paris: OECD Publishing. <https://doi.org/10.1787/f8d7880d-en>
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- RStudio Team. (2015). RStudio: Integrated Development for R. Boston MA: RStudio. <<http://www.rstudio.com/>> (27 February 2018).
- Salager-Meyer, Françoise. 1992. A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes* 11: 93–113. [https://doi.org/10.1016/S0889-4906\(05\)80002-X](https://doi.org/10.1016/S0889-4906(05)80002-X)
- Salas, Millaray D. 2015. Reflexive metadiscourse in research articles in Spanish: Variation across three disciplines (Linguistics, Economics and Medicine). *Journal of Pragmatics* 77: 20–40. <https://doi.org/10.1016/j.pragma.2014.12.006>
- Samraj, Betty. 2008. A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes* 7: 55–67. <https://doi.org/10.1016/j.jeap.2008.02.005>
- Schwach, Vera & Elken, Mari. 2018. *Å snakke fag på et språk andre forstår. Norsk fagspråk i høyere utdanning og arbeidsliv. NIFU-rapport 2018: 20*. Oslo: Nordic Institute for Studies in Innovation, Research and Education (NIFU).
- Swales, John M. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: CUP.
- Taylor, Vi Linh. 2001. Tense Usage in Academic Writing: A Cross-Disciplinary Study. Master's thesis, University of Victoria, Canada.
- de Waard, Anita & Maat, Henk Pander. 2012. Verb form indicates discourse segment type in biological research papers: Experimental evidence. *Journal of English for Academic Purposes* 11: 357–366. <https://doi.org/10.1016/j.jeap.2012.06.002>
- Wächter, Bernd & Maiworm, Friedhelm. 2014. *English-Language-Taught Degree Programmes in European Higher Education. The State of Play in 2014*. Bonn: Lemmens.

The expression of time in English and Czech children's literature

A contrastive phraseological perspective

Markéta Malá, Denisa Šebestová & Jiří Milička

Charles University, Prague

This chapter explores the expression of the concept of time in children's narrative fiction cross-linguistically, comparing Czech and English. Specifically, it analyses multi-word units and patterns which the respective languages employ when referring to time. The new Engrammer software was developed to facilitate the extraction of n-grams with lemmatised cores and positional mobility, making it possible to compare temporal patterns in English with those used in highly inflectional Czech with variable word-order. The results of the study suggest that in children's fiction in both languages, time plays an important role in structuring the text, frequently creating dramatic effects. Even though the formal means of expressing time may differ between English and Czech (e.g. diminutives in Czech vs. phrasal description in English), register appears to substantially influence the way time is framed in children's literature in both languages.

1. Language comparison, n-grams, and children's literature

The research presented in this chapter relates to three areas: first, it is a contrastive study, comparing two typologically distant languages, English and Czech. Second, the study tests the possibilities and limits of phraseological approaches to language comparison, taking as its starting point recurrent n-grams and patterns in the two languages. Third, the study adopts a register perspective on phraseology: we explore and compare the functions of the lexico-grammatical patterns in relation to a specific register – children's literature – focussing on the expression of the concept of time.

The study draws on the synchronic contrastive tradition in linguistic research. We adopt a data-driven, frequency-based approach, and set out to compare

frequent phraseological patterns in the two languages, relying on the assumption that “meaning ... reside[s] in multi-word units rather than single words” (Ebeling & Ebeling 2013: 65), such as *for the first time*, or *in the middle of*. In fiction, such “frequent sequences of words have important discourse functions as textual building blocks” (Mahlberg 2013: 56).

While multi-word units have been found to serve as a useful basis for comparing typologically related languages (Ebeling & Ebeling 2013, 2017; Hasselgård 2017), they appear rather challenging when dealing with languages which are typologically distinct, such as English and Spanish or Czech (Cortes 2008; Čermáková & Chlumská 2016, 2017; Granger 2014; Šešestová & Malá 2019). The challenges presented by the two languages compared here are due to the highly inflectional vs. predominantly analytical character of Czech and English, respectively, and the variability of word-order in the former. Both factors have some impact on the delimitation of recurrent multi-word units in Czech. For the identification of such units, the present study relies on the new software Engrammer, capable of identifying patterns with partial lemmatisation and positional mobility (see Section 4 below).

Linguistic patterning has been shown to be not only language-specific but also highly sensitive to register (e.g. Biber et al. 2004; Gries et al. 2011). The present study therefore weighs the influence of diverse language structures on the one hand, and the unifying features of the register of children’s literature on the other. The main reason why children’s literature, or more specifically, narrative children’s fiction, was selected is that it is presumed to manifest idiosyncratic features stemming from its specific functions, shaped by its target audience. The study aims to contribute to the characterisation of the register, whose linguistic analysis has received relatively little attention, despite the importance of children’s literature in “a larger socio-cultural context” (Čermáková & Chlumská 2017: 77). Since “linguistic features are always functional when considered from a register perspective ... [they] tend to occur in a register because they are particularly well suited to the purposes and situational context of the register” (Biber & Conrad 2009: 6). The key situational criterion that makes it possible to view children’s fiction as a register is the intended audience: “Children’s books are different from adults’ books: they are written for a different audience, with different skills, different needs, and different ways of reading; equally, children experience texts in ways which are often unknowable, but which many of us strongly suspect to be very rich and complex” (Hunt 2005: 3).

Linguistic analyses of children’s literature, though quite scarce (cf. Čermáková & Chlumská 2016, 2017; Hunt 2015; Thompson & Sealey 2007), have suggested that not only may specific lexical items and language structures be used in children’s fiction but that “depictions of the relationship between protagonist(s) and world [are] handled in a distinctive way compared with the depiction of that relationship in fiction aimed at adult readers” (Thompson & Sealey 2007: 4). It was shown

by Čermáková and Chlumská (2017: 173) that the major semantic categories of n-grams in children's literature comprise verbs of speaking, body-parts, time and space. Thompson and Sealey (2007: 9) ascribe the high frequency of "items relating to chronology (*time, day*)", together with items denoting people and places, to the narrative character of texts, regardless of whether they were written for child or for adult audience.¹ At the same time, however, they note that "time may be represented as having different qualities when perceived from a child's perspective. ... [it] may pass more slowly in the child's perception, and it may also be more subject to regulation by others." (Thompson & Sealey 2007: 18) Children's fiction has been observed to employ "the temporal pattern of the iterative" frequently, "most probably since the iterative reflects a child's perception of time as cyclical, non-linear, where recurrent events and routines emphasise the eternal cycle rather than the linear flow of time." (Nikolajeva 2004: 167) On the other hand, texts "based on linear time ... demonstrate the central role narrative plays in making sense of the shared human experience of being in time" (Sainsbury 2014: 189–190), and "linearity in narration ensures that the story is easy to follow" (Knowles & Malmkjaer 1996: 162). The way time is depicted seems to be closely linked to the register of narrative prose for children, and at the same time influenced by the linguistic choices available in a particular language.² Therefore, it appears interesting to explore how temporal meanings are expressed through multi-word units within the selected register, and to compare English and Czech in this respect. The identification and analysis of multi-word patterns was conducted using the Engrammer software, developed specifically to suit this type of contrastive research. Our overall aim is to compare the degree to which the multi-word units reflect register-specific features on the one hand, and language-specific typological properties of the two languages on the other.

Section 2 specifies why adverbials were chosen over other linguistic means expressing temporal meanings. Section 3 describes the corpora employed. Section 4 introduces the methodology adopted for the identification and analysis of multi-word units in our data expressing temporal meanings, and presents the specialised software which was designed to this end. Section 5 presents the results of our analysis. The patterns identified in our data are sorted by their textual

1. Wild, Kilgarrieff and Tugwell (2012: 201), however, point out differences between the expression of time and space in literature written for children and for adults: "A concern for time is striking on the adult side [...] If time is on the adult's side, then space is on the children's."

2. Cultural traditions are likely to play a role too, but our material does not make it possible to explore this factor in detail.

functions and compared cross-linguistically. Finally, in Section 6, we discuss the general implications of our study as well as suggested areas for future research.

2. Expressing time with adverbials

The present study focusses on temporal meanings expressed by adverbials rather than by the grammatical categories of the finite verb. In Czech, the verbal categories of tense and aspect are obligatory in clausal propositions; tense further serves to contextualize the verbal action in relation to the moment of speaking (Daneš et al. 1987: 87). In English, tense is an obligatory category expressed by finite verb forms (Quirk et al. 1985: 149). Temporal adverbials, on the other hand, are largely optional elements in both the languages. Their distribution, however, is related to verb semantics (e.g. individual vs. stage level predicates) and verbal tense and aspect (Karlík 2017). Both in Czech and in English, temporal adverbials have various realisation forms: noun phrases (e.g. *tento večer* / *this evening*),³ prepositional phrases (e.g. *o Vánocích* / *at Christmas*), adverb phrases (e.g. *nakonec* / *finally*), or clauses (e.g. *až přijedete* / *when you arrive*). In each of these forms, the temporal meaning is signalled by a temporal expression – a noun (in noun and prepositional phrases), an adverb, or a temporal conjunction. Even though temporal relations may be expressed by prepositions, the extent of their polysemy makes it impossible to rely on prepositions as indicators of temporal meanings (cf. Hasselgård 2017: 82; Karlík 2017). This can be illustrated by English examples such as *on Monday* (temporal), *on the floor* (spatial); similarly in Czech: *v pondělí* ('on Monday', lit. 'in Monday'), *v pokoji* ('in the room'). For this reason, prepositions will not be used as the basis for our temporal adverbial queries.

In addition to temporal adverbials, time-related meanings can be conveyed by noun phrases with a temporal head which serve as nominal clause elements, e.g. the subject in *Our time has come*.⁴ These syntactic functions, however, do not appear to be performed by patterns based on frequent n-grams explored in this study.

3. The corpora

The material is drawn from two broadly comparable corpora of children's fiction which were compiled on the basis of the same criteria (text type = fiction,

3. The Czech – English pairs of phrases in this sentence have the same meaning.

4. All the examples come from the two corpora we worked with, *BNC-Jun* and *SYN-7-Jun* (see Section 3).

audience = child + teenager), as sub-corpora of the *British National Corpus* and the Czech general corpus *SYN-7*. The size of the corpora is given in Table 1. The difference in the number of texts is due to the fact that while the *BNC* consists of text samples, *SYN-7* comprises full texts. We worked with lemmatised versions. Both corpora were limited to original texts, i.e. excluding translations. The broad category of texts written for child and teenage audience was narrowed down in both languages to fiction (excluding academic prose, non-academic prose and biography, and other published and unpublished written material).

Admittedly, the publication time spans of texts included in the two corpora do not correspond, although they do overlap to a degree. While the fiction component of the *BNC* contains texts published between 1960 and 1994 (with 75% published after 1975),⁵ the Czech *SYN-7* corpus spans the period between 1967–2014.⁶ Still we have opted for these two corpora, mainly for the following two reasons: they were readily available and extensive; and they were designed as balanced and representative; to date, no English corpora of more recently published children's fiction are freely available (to our knowledge).

Table 1. The corpora used in this study

Language	Corpus	Subcorpus of	Size – tokens	Size – texts	Text publication dates span	Number of authors represented
English	<i>BNC-jun</i>	<i>BNC</i> (1994)	2,046,755	76 (excerpts)	1960–1994	36 + 29 adapted classics*
Czech	<i>SYN-7-jun</i>	<i>SYN-7</i> (2010)	2,821,044	59 (whole books)	1967–2013	43

*Oxford Bookworms edition – adapted texts.

4. The method: From frequent temporal expressions via n-grams to temporal patterns

4.1 Frequency lists

As pointed out by Stubbs (2007: 100), “many phrases are frequent because they are conventional ways of expressing common meanings.” Since we aimed at exploring

5. *BNC User Reference Guide* <<http://www.natcorp.ox.ac.uk/docs/URG/BNCdes.html#BNCcompo>> (18 March 2020).

6. *Corpus SYN version 7* <<https://wiki.korpus.cz/doku.php/en:cnk:syn:verze7>> (18 March 2020).

how the concept of time is commonly expressed in children’s narrative fiction, we decided to focus on frequent multi-word temporal expressions. In our study, these are n-grams which contain a ‘temporal core’, i.e. a lexical word with temporal meaning (e.g. *she had never, v tu chvíli se* (‘at that moment *se*-reflexive’)) or a temporal conjunction (e.g. *and when he, a když se* (‘and when *se*-reflexive’)). Our starting point was a lemma frequency list compiled for both the English and the Czech corpus. Temporal lemmata,⁷ such as *when, time, never; pak, den, když* (‘then, day, when’), were selected from the most frequent 500 lemmata in each list, resulting in 28 items for English and 33 for Czech (listed by frequency in Table 2). These lemmata were then used as ‘cores’ of n-grams with positional mobility. Interestingly, many of these temporal lemmata correspond across the two languages, although the respective frequencies of these translation equivalents differ (cf. *then / pak, when / když, moment / chvíle*, correspondences are marked in bold in Table 2). For the detailed analysis (see Section 5), the set of patterns was further restricted to those with noun cores.

Table 2. The most frequent time-related lemmata in the two corpora

	Most frequent temporal lemmata
English (BNC-jun)	<i>then, when, now, time, again, day, never, night, year, once, while, ever, always, moment, suddenly, soon, until, morning, minute, later, hour, week, already, evening, sometimes, late, ago, often</i>
Czech (SYN-7-jun)	<i>už (‘already’), když (‘when-relative’), pak (‘then’), teď (‘now’), den (‘day’), chvíle (‘while/moment’), hned, brzy (‘soon/early’), rok (‘year’), nikdy (‘never’), pořád (‘always/still/all the time’), čas (‘time’), potom (‘then’), znovu (‘again’), noc (‘night’), jednou (‘once’), dlouho (‘long / for a long time’), najednou (‘suddenly’), ráno (‘morning’), hodina (‘hour’), doba (‘time’), dnes (‘today’), kdy (‘when-interrogative/relative’), večer (‘evening’), konečně (‘finally’), stále (‘still’), vždycky (‘always / every time’), nakonec (‘finally’), již (‘already’), teprve (‘only’), někdy (‘sometimes’), zatím (‘meanwhile’), zítra (‘tomorrow’)</i>

4.2 N-grams

For both languages, n-grams were defined as “recurring strings, with or without linguistic integrity” (Lindquist & Levin 2008: 144). We focussed on 3-grams and 4-grams comprising the temporal ‘core’ at any position within the n-gram. Both

7. In the present study, we excluded lemmata which frequently refer both to time and another semantic field, such as *end*, which occurs in temporal as well as spatial expressions. A major drawback of this decision is that it narrows down the scope of the study and makes direct comparison with the related study by Čermáková & Chlumská (2016) more complicated.

these features, viz. comprising a temporal lemma (selected over word-form due to the morphological richness of Czech nominal paradigms) and allowing for variable position of the lemma in the n-gram (required by highly flexible word order in Czech), expand the range of temporal multi-word expressions identified in the corpora and facilitate English-Czech comparisons. The lemma DAY, for instance, can occur in the singular or plural, in the initial, medial or final position within the n-gram (*day and night, one day when, in those days*). The advantages of searching for n-grams comprising a specific mobile lemma are more obvious in Czech. For example, the lemma DEN ('day') may vary in number, case and position within the n-grams: *den a noc* ('day and night'), *tři dny a tři* ('three days and three'), *ve stejný den* ('on the same day'); and Czech 3-grams which differ merely in word-order can be grouped together: *jednoho dne se* ('one day se-reflexive'), *se jednoho dne* ('se-reflexive one day').

The choice of lemmata over word-forms as our first step towards the identification of patterns is motivated by our effort to identify as many patterns involving the given temporal expression as possible. Without lemmatisation, some low-frequency patterns would likely go unnoticed (particularly in Czech, due to its morphological variability). A closer examination of the patterns reveals that typically, a pattern with a specific function contains only a particular word-form (rather than the full lemma). This is true for both English and Czech patterns (e.g. *day by day* rather than **days by days*; *hodnou chvíli* – singular 'a nice moment' rather than **hodné chvíle* – plural 'nice moments'⁸). This finding supports the notion that our identified chunks are indeed patterns, phraseological units, rather than free word combinations, as they are (to a degree) fixed. This is in line with Sinclair's (2004: 31) observation that the collocates of a singular word form may not overlap with those of its plural variant (*blue eyes* rather than *blue eye*; *in your mind's eye* not **in your mind's eyes*).⁹

The length of the n-grams explored in this study, 3 to 4 words, is determined by the focus on semantic classification: 2-grams are difficult to classify semantically since they comprise mostly grammatical words; moreover, "shorter bundles are often incorporated into more than one longer lexical bundle" (Biber et al. 1999: 990). On the other hand, 5-grams and longer multi-word units are often semantically complex. Since in Czech subordinators are obligatorily preceded by a

8. Interestingly, in case of *hezkou chvíli* ('a pretty long time', lit. 'a nice moment'), replacing the singular word form with plural would change the meaning of the noun phrase and result in a loss of idiomaticity (the plural *hezké chvíle* has the literal meaning of 'nice moments' and cannot be used as a time adverbial).

9. We would like to thank the anonymous reviewer for suggesting the Sinclair reference.

comma, e.g. *chvíle, kdy jsme* ('moment when we'), we decided to include punctuation in the n-grams. N-grams containing commas seem to capture the form of the temporal expressions in Czech more realistically, while those containing quotation marks highlighted the high frequency of direct speech in children's fiction in both languages and its role in structuring the narrative, e.g. "...*Somebody should stop him!*" *Just then Martha ran into the room.*

4.3 The tool used: Engrammer

The n-grams comprising the selected temporal lemmata were identified using *Engrammer*, a newly developed tool dedicated to n-gram and skipgram¹⁰ extraction and analysis. The tool is freely available from <http://milicka.cz/en/engrammer>, and can be used offline with the user's own data. Engrammer is designed to accommodate a range of typologically different languages.

The tool extracts n-grams allowing for numerous customisations and experimenting with various research designs, e.g. using lemmatization or the possibility to explore unordered n-grams (i.e. n-grams with positional mobility).¹¹ These two options were particularly relevant to our research (cf. 4.2).

The main advantage of Engrammer over standard corpus managers is that it retrieves not only the most frequent n-grams that contain a certain word form/lemma, but also ranks the n-grams according to the degree of association with the given word form/lemma, thus making it possible to find the most typical phrases of that word form/lemma. The ranking is performed according to an association metric which can also be specified by the user. The primary association metric we used is the lower limit of the confidence interval of the Risk Ratio (also known as Relative Risk, cf. Daly 1998).¹²

Let us illustrate the interpretation of the metric by the following example: imagine a 2-million-word corpus in which the word *time* occurs 3,378 times and the 3-gram *long time ago* 24 times. The combination of *long * ago* (the asterisk representing any given word) occurs 26 times. Our metric determines that if the

10. Skipgrams are defined as possibly discontinuous sequences of elements, i.e. sequences that may contain gaps (hence they are also called 'gappy n-grams'). The notion of the skipgram was probably introduced into linguistics and NLP by Navrátil & Zuhlke (1997), but the idea is even older, cf. Schukat-Talamazzini et al. (1995).

11. Technically speaking, these structures are not n-grams, since n-grams are by definition sequences. It would be appropriate to refer to them as 'word multisets'. However, as the multiset theory is not widely known, we have decided to call the structures 'unordered n-grams'.

12. We are using the rather infelicitous term 'Risk Ratio (of n-grams)' in a technical sense, with no negative connotations intended.

Table 3. Contingency table of absolute frequencies of the word *time*

This metric also allows for ranking n-grams of various lengths together (i.e. the n-grams most strongly associated with the given search term will be displayed, irrespective of their length). To avoid an excessive number of similar results (e.g. *long time ago*, *a long time ago* and *be a long time ago* – note that this search is lemmatised), the program thus makes it possible to hide overlapping and very similar variants automatically, and to show them only at the request of the user.

Figure 1 illustrates the graphic interface of Engrammer. The first column represents the retrieved n-grams. After clicking on the n-gram in the first column, the interface displays: (1) the hidden variants of the chosen n-gram (in the second column, e.g. *long time ago*, *be a long time ago*); (2) the words alternating in the position occupied by the selected search term (in the third column, e.g. *long time ago* – *long year[s] ago*); (3) a KWIC view of the n-gram instances is shown (in the rightmost panel).



4.4 Temporal patterns

On the basis of n-grams strongly associated with the selected temporal lemmata, recurrent temporal patterns were identified in both languages. Following Lindquist & Levin (2008: 144), we use the term ‘pattern’ for “meaningful, linguistically structured recurring sequences of words”. A qualitative analysis of the patterns in context revealed the textual functions of the patterns in children’s narrative fiction in the two languages. The temporal patterns, their functions and cross-linguistic correspondences will be dealt with in Section 5.

4.5 An illustration of the method: *time* and *doba*

To illustrate the method, we can take the lemma *TIME* as our starting point. The lemma *TIME* occurs in various n-grams; it is found most frequently (125 times) within the 4-gram *for the first time*. This 4-gram is strongly associated not only with the word *time* (out of the 144 occurrences of the 4-gram *for the first **, only 19 comprise an expression other than *time*), but also generally with expressing temporal meanings. The remaining 19 instances of the 4-gram *for the first ** are also overwhelmingly temporal, containing nominal expressions that carry temporal meanings, e.g. *couple of weeks, few years, six months, hour, week, night, two days*, etc. Expanding the context in which the n-gram *for the first time* occurs, the concordance lines (Figure 2) reveal a temporal pattern *for the first time in* ‘a long period of time’, associated with intensification and with signalling a change.

when the way ahead seemed ,	for the first time	in ages , to be solid under her feet , Marie
door , which opened , slowly ,	for the first time	in ten years . She walked quickly in and shut
that grew at the end of the lane .	For the first time	in days his eyes seemed to acknowledge her .
with the Duke 's body and ,	for the first time	in her life , she entered Buckingham Palace .
and anything else he needed .	For the first time	in all his journeys he found a room that was
And that was a surprise .	For the first time	in her life she 'd allowed someone to hook on
needles through a frozen limb .	For the first time	in her life , perhaps , the Astropath had been
I was in great pain , and suddenly	for the first time	in my life , I forgot my fear of John Reed
actually looked at a flower , and	for the first time	in ten years he realized how beautiful
demented birds , disturbed	for the first time	in who knows how long , were battering
to hear that Matthew was dead .	For the first time	in his life , Matthew Cuthbert was an
her business on a sound footing	for the first time	in her life . ` It 'd be a grand thing
the elements as slaves , then	for the first time	in history slavery will be abolished . Human
a tankard . Blake was content	for the first time	in days . The barman returned , smiled , and
old man , propped in his chair	for the first time	in over a month , laid a trembling hand on the

Figure 2. The temporal pattern *for the first time in* ‘a long period of time’

The same procedure may be applied in Czech. The 3/4-grams closely associated with the lemma *DOBA* (‘time, period’) were found to comprise either the demonstrative

pronoun *ta* (*té, tu* ('that-GEN', 'that-ACC'): *od té doby (se), od té doby, (co/než), (po) celou tu dobu, (. /ale) do té doby* ('(and) since that time (*se*-reflexive), since the time (when), (for) all that time, (. /but) till that time')), or the indefinite pronoun *nějaká* (*nějaké, nějakou* ('some-LOC', 'some-ACC'): *(.) po nějaké době, za nějakou dobu* ('(.) after some time, in some time').¹³ In Czech, the expression of noun definiteness is not obligatory. The demonstrative pronouns can be used to mark endo- or exophoric contextual ties explicitly. The colligation pattern '*ta + doba*' is slightly more frequent in children's than in general fiction.¹⁴ This may suggest the importance of explicit anaphoric ties in the structure of children's narrative fiction as well as the summarising effect of the temporal pattern with the demonstrative (Examples 1 a, b).

- (1) a. Po léta dbáme o jeho výchovu, měl nejlepší učitele, ale výsledek je hrozný. Za celou tu dobu se nenaučil téměř nic.
 ['For years we have taken care of his education, [...], but the result is terrible. In all that time he has barely learnt anything.']
 b. Od té doby se ohromný rudý kámen jmenuje Morinova skála...
 ['Since that time the huge red rock is called Morin's Rock...']

The patterns with the indefinite pronoun, on the other hand, refer to periods of time of indefinite (not short) duration (Example 2).¹⁵

- (2) Po nějaké době přišel do královského města a tam uviděl krásný palác.
 ['After some time he came to the royal town and saw a beautiful palace there.']

To summarise this section, the process of identifying temporal patterns adopted in this study is illustrated in Figure 3.

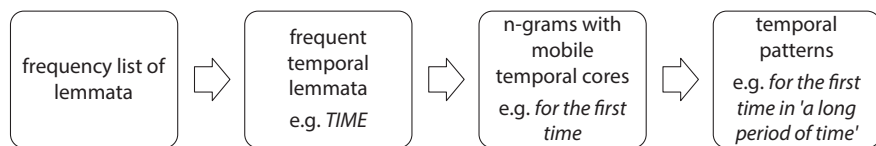


Figure 3. Identification of temporal patterns

13. Overlapping n-grams are merged in these examples of 3/4-grams with *doba*.

14. 38% of noun phrases headed by the lemma *doba* include the demonstrative *ta* in children's fiction, while in general fiction (SYN-7) the proportion is 31%.

15. The noun *doba* actually always seems to require some modification (either pre- or postmodification), as attested both in children's and in general fiction corpora. We thank an anonymous reviewer for pointing this out.

5. Results and discussion

Once textual patterns were identified in both our English and Czech datasets, we analysed them qualitatively, examining their contexts to determine their textual functions. These functions are discussed in Section 5.1.

There appears to be a major difference between patterns with noun cores, and those with adverbs or conjunctions as the cores in expressing the textual functions. The noun patterns typically comprise more than one temporal expression, and the temporal function is associated with the pattern as a whole, e.g. patterns with DAY included *day and night; hours a day; one day when; day after day, every day, three days later*. In most n-grams with adverb or conjunction cores, on the other hand, the temporal meaning stemmed merely from the core adverb/conjunction (e.g. *I've **always** wanted; **when** she arrived*). When identifying the textual functions, we focussed first on noun temporal patterns. A quantitative summary of functions of patterns based on selected nouns (namely the nouns which have translation counterparts represented in both the English and Czech data-sets) is presented in Table 6. We then also explored the functions of recurrent temporal patterns with adverbs and conjunctions. These, however, were not quantified, and will be used in Section 5.1 to give a more comprehensive view of the means of expressing the textual functions.

Section 5.2 compares the patterns containing equivalent lexical cores between the English and Czech data, focussing on the impact of typological differences. To ensure cross-linguistic comparability, we decided to limit the comparison to those temporal nouns whose equivalents were represented in both languages. The resulting set of nouns (8 per each language) which served as the basis for the comparison is presented in Table 4. (Some nouns have more than 1 equivalent in the other language.)

Table 4. Correspondence of temporal nouns in the data-set

English cores	Czech equivalent cores
<i>day</i>	<i>den</i>
<i>moment / while</i>	<i>chvíle</i>
<i>year</i>	<i>rok</i>
<i>night</i>	<i>noc / večer</i>
<i>time</i>	<i>čas / doba</i>
<i>morning</i>	<i>ráno</i>
<i>evening</i>	<i>večer</i>

We established an arbitrary cut-off point at the 'Risk Ratio of n-gram' value (collocation strength metric) of 50 at least. The frequencies of patterns including the noun lemmata which have direct translation counterparts in the dataset are shown in Table 5.

Table 5. Pattern type frequencies for English and Czech

ENGLISH	TIME	MOMENT	WHILE	DAY	YEAR	NIGHT	MORNING	EVENING	TOTAL
TOTAL	232	105	55	170	144	134	56	45	941
RISK* \geq 50	104	52	20	69	88	61	30	20	444
CZECH	ČAS	DOBA	CHVÍLE	DEN	ROK	NOC	RÁNO	VEČER	TOTAL
	<div> <div>'TIME'</div> <div>'MOMENT'/'WHILE'</div> <div>'DAY'</div> <div>'YEAR'</div> <div>'NIGHT'</div> <div>'MORNING'</div> <div>'EVENING'</div> </div>								
TOTAL	91	76	256	242	143	74	50	36	968
RISK* \geq 50	45	60	104	111	85	25	27	14	471

*'Risk Ratio of n-gram' value

5.1 Textual functions of temporal patterns

In this subsection we will outline the functional classification of temporal patterns, which applies both to the Czech and English data, i.e. patterns fulfilling these functions were attested in both languages. We have not identified any functions exclusive to one language only; however, individual functions may manifest themselves in different ways in each language. The lack of functions clearly limited to one language suggests that these functions may be shared by texts intended for children regardless of language, possibly determined by the specific register characteristics. However, research into comparable corpora of adult fiction is needed to verify this. The cross-linguistic differences seem to be grounded typologically and potentially also culturally; they will be discussed in Section 5.2.

During the functional classification of patterns, we decided to assign only one (dominant) textual function to each pattern, although we are aware that some of the n-grams may perform various functions depending on the context, or "a single occurrence" of an n-gram "can be considered multifunctional" (Biber 2006: 139).

The patterns we have identified fulfil the following major narrative-organising functions:

- explicit signalling of time (*at that moment*)
- intensification (*day after day*)
- signposting a dramatic moment or change in the narrative (*just in time to*)
- marking vague time (*once upon a time*).

Another factor, which overlaps with the textual functions, is idiomaticity. This was understood in broad terms, to encompass invariable non-compositional patterns which carry non-literal meaning.

Table 6 shows the numbers of pattern types fulfilling each function (within the limited sample of patterns with ‘Risk Ratio of n-grams’ value ≥ 50), as well as the number of idiomatic patterns. Some patterns were not found to fulfil any of these functions – their numbers are listed under ‘Other’.

Table 6. Type frequency of textual functions of patterns (‘Risk Ratio of n-grams’ ≥ 50); note that idiomaticity is an additional feature, overlapping with the textual functions

Function	TIME	MOMENT	WHILE	DAY	YEAR	NIGHT	MORNING	EVENING	Total
Explicit time	44	4	0	32	67	39	23	11	220
Intensification	31	7	2	8	10	8	1	1	68
Dramatic	4	38	12	0	0	0	1	0	55
Vague time	7	2	1	26	0	1	1	3	41
Other	18	1	5	3	11	13	4	5	60
Total	104	52	20	69	88	61	30	20	444
<i>Idiomaticity</i>	4	2	1	4	3	4	1	1	20

Function	ČAS	DOBA	CHVÍLE	DEN	ROK	NOC	RÁNO	VEČER	Total
	‘TIME’	‘TIME’	‘WHILE’/ ‘MOMENT’	‘DAY’	‘YEAR’	‘NIGHT’	‘MORNING’	‘EVENING’	
Explicit time	1	27	25	62	59	9	13	3	199
Intensification	9	15	20	20	13	9	7	5	98
Vague time	13	0	47	18	7	3	0	0	88
Dramatic	8	0	12	0	0	0	0	0	20
No function	14	18	0	11	6	4	7	6	66
Total	45	60	104	111	85	25	27	14	471
<i>Idiomaticity</i>	8	1	3	8	2	4	4	3	33

The individual functions will now be discussed and illustrated by examples, ordered by their frequency in the English data. To give a more comprehensive picture of the expression of the particular functions, the examples also include temporal patterns with cores which are not included in the quantitative summary in Table 6 (i.e. other nominal and non-nominal cores).

– Explicit signalling of time

Various aspects of temporal relations are made explicit by multi-word patterns:

- a. Specific temporal reference to a moment or punctual action, which may be repeated, as in *three times a day*; *v tu chvíli* ('at that moment'). The "magical number" three occurs fairly often.
- (3) ...a žili tam šťastně a spokojeně. Každého rána přicházely k jejich chatrči děti nebo nějaká stařenka a přinášely jim smažené placičky.
[‘...and they lived there happily and in peace. Every morning some children or an old woman would come to their shed and bring them pancakes.’]
- b. Alternatively, there is lexical explicitation of the duration of an action, helping to structure the narrative and/or build suspense. In example (4), for instance, the sequence of setting out on a journey and coming to the town would be obvious without the adverbial; the expression *po nějaké době* ('after some time') marks the (limited) duration of the journey and the onset of a new stage in the story. Examples include *for a moment*, *for a little while*, *for a long time*; *po nějakém čase*, *hodnou/hezkou chvíli*, *čas plynul*, *tři dny a tři noci* ('after some time, a pretty long time, the time passed, three days and three nights').
- (4) Princ mu dal svůj ostrý, silný meč a tesař se vydal na cestu. Po nějaké době přišel do královského města a tam uviděl krásný palác.
[‘The prince gave him his sharp, strong sword and the carpenter set out on his journey. After some time, he came to the royal town and saw a beautiful palace.’]
- c. Other examples contain explicit marking of temporal relations between two or more actions, happening simultaneously or in sequence: *the next morning*, *but after a while*, *for the first time*, *at that moment*; *at the same time*, *after a moment's hesitation*; *po nějakém čase* ('after some time'), *nas-tal čas* ('time has come'), *tak – a teď* ('so – and now'), *když nadešel čas* ('when time came'), *jednoho dne* ('one day'), *teprve teď* ('only now'), *hned zítra* ('right tomorrow').
- (5) And he passed me a pistol. At the same time he began to move quietly and, after a few steps, the hole was between us and the other five.

– Intensification

The patterns associated with intensification refer to the passing of time or repetition. In both languages, typical intensifying patterns contain parallel adverbial structures consisting of noun or adverbial phrases linked by coordinators or prepositions, often repeated, such as *day after day*, *day by day*, *once and for all*, *day and night*, *again and again*, *over and over again*; *ve dne v noci* ('day and night'), *večer co*

večer ('evening by evening'), *čas od času* ('from time to time'), *den ze dne* ('day by day'), see example (6).

- (6) Night and day. Day and night I can hear that wheel turning and the machines going.

– **Signposting a dramatic moment or change in the narrative**

As observed by Thompson and Sealey (2007), multi-word patterns are used in children's fiction to mark a dramatic turning point in the narrative. These patterns typically contain or co-occur with intensifiers or focalisers, such as *just*, *only*, *právě* ('just'): *just in time to*, *then/when suddenly*, *at this very moment*, *at that moment* (*there appeared*), *too late*; *právě v tu chvíli* ('just at that moment'), *právě ve chvíli* (, *kdy*) ('just at the moment (when)'), *nemáme moc času* ('we don't have much time'), *(ne)ztrácet čas* (('not) to lose time'), *každou chvíli* ('every now and then / any minute now'), *je nejvyšší čas* ('it is high time'), *v poslední chvíli* ('at the last moment'). Examples (7) and (8) provide wider contexts to illustrate the focalising function of these patterns in the text.

- (7) Mary gave one horrified glance at it, then flung herself madly into the ditch at the side of the lane. She was only just in time to escape being knocked down.
- (8) Rytíř stačil jen v poslední chvíli uskočit a odrazit vidličku mečem.
[At the last moment the knight just managed to jump aside and fend off the fork with his sword.]

– **Vague time**

A number of patterns express vague temporal reference, typically to the past. The indefiniteness of temporal reference contributes to the construal of the narrative as unreal or supernatural. "[F]airy tale writers may facilitate our perception of our own world as magical by refraining from explicit spatiotemporal staging, or by providing an impression of temporal and spatial distance between reader (and writer) and story" (Knowles & Malmkjaer 1996: 160, emphasis is ours).

Examples include *once upon a time*, *in those days*, *a long time ago*; *one evening*, *one day when*, *many years ago*; *před dávnými časy/věky/léty/lety* ('many ages/years ago'), *I* ('in times long past'), *jednoho dne* ('one day'). Some of these patterns are register-specific (*once upon a time*¹⁶), or their usage may be evocative of fairy stories (*před dávnými léty* ('long time ago')).

16. *Once upon a time* occurred outside our analysed sample (Risk Ratio of n-gram = 40).

- (9) Před dávnými časy žil v jedné vesnici v Kašmíru mudrc, ...
 ['Long time ago in a village in Kashmir there lived a wise man...']

– Idiomatic expressions

Idiomatic patterns have no single function; rather, they seem to play the role of style markers, aimed towards fulfilling the aesthetic function of the text. Likewise, they may contribute to another specific function of children's fiction, namely the educational one (cf. Čermáková & Chlumská 2016), providing the young readers with varied lexical input so as to expand their vocabulary. This is a heterogeneous group represented mostly by patterns with adverbial functions, e.g. *day and night*, *spur of the moment*, *from time to time*, *once in a while*, *hang on a minute*; 'den ze dne + comparative' ('day by day + comparative'), *hodnou/hezkou chvíli* ('a pretty long time', lit. 'a nice moment'), *do roka a do dne* ('in a year and a day's time'), *dočkej času jako husa klasu* (proverb roughly equivalent to 'Rome wasn't built in a day', lit. 'wait for your time like a goose does for her ear of wheat'), *do nejdelší smrti* ('forever and a day', lit. 'until the longest of deaths'), *od rána do večera* ('from morning to evening').

5.2 A contrastive look at temporal patterns

As mentioned in Section 5.1, idiomaticity seems typical of the register of children's literature. In English, idiomatic patterns comprise 4 per cent of patterns with noun cores in our sample (20 out of 444 patterns), and in Czech 7 per cent of the patterns with noun cores (33 out of 471 patterns). This may suggest that our Czech texts are more idiomatic overall; the difference, however, is not statistically significant.

In Czech, some of the idiomatic pattern types overlap, forming longer patterns (*dočkat času jako + času jako husa = dočkat času jako husa klasu* ('wait for your time like a goose does for her ear of wheat')). Other types can be considered variations of one pattern (*právě/zrovna v tu/tuto chvíli* ('just at that/this moment'); or *být/mít nejvyšší čas, aby* ('be/have high time to')). By contrast, for English idiomatic patterns, no such overlaps between pattern types were identified.

Overall, the examples mentioned in Section 5.1 were found to be more frequent in children's books than in fiction written for adult audience. Table 7 compares the frequency of the most frequent idiomatic patterns in children's and adults' fiction subcorpora of SYN-7 and BNC;¹⁷ the absolute frequencies, however, are too low to be conclusive, and the quantitative results should be accompanied

17. The sizes of the children's fiction subcorpora are given in Table 1; those of the adults' fiction subcorpora are 33,047,774 tokens (780 texts) for SYN-7 and 17,317,696 tokens (376 texts) for BNC.

by a qualitative functional analysis of the phrases in adults' fiction before any conclusions can be drawn.

Table 7. The frequency of idiomatic temporal noun patterns in children's and adults' fiction (ipm = instances per million tokens).

English (BNC)	Children's fiction		Adults' fiction	
	Total hits	ipm	Total hits	ipm
just in time	22	10.749	144	8.315
time to time	19	9.283	308	17.785
day and night	15	7.329	60	3.465
in no time	10	4.886	93	5.370
all day long	9	4.397	34	1.963
day after day	8	3.909	59	3.407

Czech (SYN-7)	Children's fiction		Adults' fiction	
	Total hits	ipm	Total hits	ipm
od rána do večera	33	11.7	231	6.99
čas od času	31	10.99	580	17.55
nejvyšší čas	30	10.63	293	8.87
dnem i nocí	14	4.96	58	1.76
právě v tu chvíli	8	2.84	32	0.97

Other differences between English and Czech patterns are grounded in the typological characteristics of the languages compared. The most salient differences are related to non-correspondences in n-gram length. English, as a predominantly analytical language, employs articles, which causes noun-based n-grams to be longer than their Czech equivalents (*za chvíli* / *after a while*). Other length non-correspondences are caused by the fact that Czech, being synthetic, prefers to express grammatical meanings as well as some lexical modifications by suffixes, e.g. a Czech case ending in lieu of an English preposition (*chvíli/chvilku* ('moment-accusative') – *for a moment, for a (little) while*). Similarly, the diminutive meaning is expressed by a diminutive suffix in Czech, corresponding to an English adjective (*za chvíličku* ('after while-diminutive') – *a little while later*).

Some differences can also be observed in the composition of n-grams. N-grams of comparable length may comprise a personal pronoun in English, while their Czech equivalents contain a verb form, or the pronoun *se*, a constituent of reflexive verbs in Czech. The Czech verb expresses the grammatical categories of number and person in its personal ending: *za/po chvíli byli/bylo/se/řekl* ('after

while they-were/it-was/ *se*-reflexive/he-said'). In English, the categorial meanings of number and person are expressed separately through personal pronouns, *after a while* (*he/she/I/the*), which are typically followed by a lexical verb, as in *after a while he stopped/said/began*. Thus, we are ultimately dealing with another case of n-gram length non-correspondence.

Another type of typologically based difference consists in word order within the n-grams. Apart from the positional variability in Czech, as mentioned in Section 4, some patterns illustrate the different positions of modifiers. In *právě/zrovna v té/tu chvíli* ('exactly/just at that moment'), the focussing adverbs *právě* or *zrovna* premodify the whole prepositional phrase *v té/tu chvíli*. These patterns correspond to the English *at this/that very moment*, where the intensifier *very* premodifies only the head noun. Different dependency relations in the noun phrase may cause word order differences, as in the pattern *po chvíli ticha* ('after a moment silence-genitive') corresponding to *a moment's silence*.

To sum up, the similarities between the patterns in both languages appear to stem from their textual functions, associated with the register of children's narrative fiction. The differences can be ascribed to linguistic differences between English and Czech, and – in the case of idiomaticity – perhaps to the culture-specific features of the register.

6. Conclusions

Our methodological question concerned the potential assets and pitfalls of employing n-gram extraction in contrastive analysis. The study suggests that n-grams can be used as a convenient stepping stone towards the identification of patterns, allowing for automatic extraction of a large amount of data. We started from a frequency list aiming to identify frequent temporal expressions, subsequently used these as the basis for n-gram extraction, and proceeded towards temporal patterns with the help of qualitative analysis of the n-grams in context.

The Engrammer software, allowing for lemmatisation, grouping n-grams of varying lengths, and the extraction of unordered n-grams, helps overcome some of the caveats stemming from the typological differences between English and Czech (non-correspondences in n-gram length and word order). Still, the application of n-gram-based approaches to inflectional languages like Czech has its limitations. A contrastive approach offers a suitable complementation to the n-gram method, revealing the ways in which n-grams are influenced by typological characteristics of the respective languages (Section 5.2).

Within the core lemmata which were used as the basis for the identification of patterns, there seems to be a difference between individual word classes. In

patterns based on temporal adverbs or conjunctions, the temporal meaning tends to be limited to the core adverb/conjunction only, e.g. *'ve always wanted, always too afraid, when he saw me*. We have identified few typical temporal patterns with specific textual functions here (e.g. *then/when suddenly I*). On the other hand, nouns occur in patterns with specific temporal functions which are often not derivable from the noun itself. Using Engrammer also enabled us to identify temporal constructions in which different temporal nouns may alternate in a given slot, e.g. *for the first time/couple of weeks/six months; day/morning and night; hours a day/month/week; in those days/months/weeks*.

The study has suggested that time in children's fiction is organised in specific ways (Section 5.1). Further, children's fiction seems to be marked by a fair degree of idiomaticity: some of the idiomatic patterns are register-specific (*once upon a time*). We suggest that the idiomatic language may be used purposely with regard to the intended audience of the texts and their specific needs. As children's literature fulfils an aesthetic as well as an educational role (Čermáková & Chlumská 2016: 163), the authors may be employing rich idiomatic language in order to attract the readers' attention as well as to help develop their vocabulary. Admittedly, a possible caveat lies in the influence of idiosyncratic authorial idiolects. For example, our Czech children's fiction subcorpus contained 4 instances of the idiomatic pattern *to BÝT doba, než* ('it TAKE a long time before', lit. 'that BE a time before'), all from the same source text *Školák Kája Mařík* by Felix Háj. In order to compensate for this potential effect, a larger and more diverse corpus would be needed.

Finally, the scope of the research could be expanded by conducting a similar study on parallel rather than comparable corpora, examining Czech-English translation equivalents of temporal n-grams. Another worthwhile research question would be to what extent the patterns identified in our data occur in adults' fiction and whether they perform similar functions there.

References

- Biber, Douglas. 2006. *University Language. A Corpus-based Study of University Registers* [Studies in Corpus Linguistics 23]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.23>
- Biber, Douglas, Conrad, Susan & Cortes, Viviana. 2004. If you look at ... Lexical bundles in university lectures and textbooks. *Applied Linguistics* 25: 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, Douglas & Conrad, Susan. 2009. *Register, Genre and Style*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511814358>
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Čermáková, Anna & Chlumská, Lucie. 2016. Jazyk dětské literatury: kontrastivní srovnání angličtiny a češtiny. In *Jazykové paralely*, Anna Čermáková, Lucie Chlumská & Markéta Malá (eds), 162–187. Prague: NLN.

- Čermáková, Anna & Chlumská, Lucie. 2017. Expressing PLACE in children's literature. Testing the limits of the n-gram method in contrastive linguistics. In *Cross-linguistic Correspondences: From Lexis to Genre* [Studies in Language Companion Series 191], Thomas Egan & Hildegunn Dirdal (eds), 75–95. Amsterdam: John Benjamins.
<https://doi.org/10.1075/slcs.191.03cer>
- Cortes, Viviana 2008. A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora* 3(1): 43–57. <https://doi.org/10.3366/E1749503208000063>
- Daly, Leslie E. 1998. Confidence limits made easy: Interval estimation using a substitution method. *American Journal of Epidemiology* 147(8): 783–790.
<https://doi.org/10.1093/oxfordjournals.aje.a009523>
- Daneš, František, Hlavsa, Zdeněk, Grepl, Miroslav et al. 1987. *Mluvnice češtiny 3. Skladba*. Prague: Academia.
- Ebeling, Jarle & Ebeling, Signe Oksefjell. 2013. *Patterns in Contrast* [Studies in Corpus Linguistics 58]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.58>
- Ebeling, Signe Oksefjell & Ebeling, Jarle. 2017. A cross-linguistic comparison of recurrent word combinations in a comparable corpus of English and Norwegian fiction. In *Contrasting English and other Languages through Corpora*, Markéta Janebová, Ekaterina Lapshinova-Koltunski & Michaela Martinková (eds), 2–31. Newcastle upon Tyne: Cambridge Scholars.
- Granger, Sylviane. 2014. A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast* 14(1): 58–72. <https://doi.org/10.1075/lic.14.1.04gra>
- Gries, Stefan T., Newman, John & Shaoul, Cyrus. 2011. N-grams and the Clustering of Registers. *The ELR Journal* 5. <<http://ejournals.org.uk/ELR/article/2011/1>> (11 May 2019).
- Hasselgård, Hilde. 2017. Temporal expressions in English and Norwegian. In *Contrasting English and other Languages through Corpora*, Markéta Janebová, Ekaterina Lapshinova-Koltunski & Michaela Martinková (eds), 75–101. Newcastle upon Tyne: Cambridge Scholars.
- Hunt, Peter. 2005. *Understanding Children's Literature*. London: Routledge.
- Hunt, Sally. 2015. Representations of gender and agency in the *Harry Potter* Series. In *Corpora and Discourse Studies. Integrating Discourse and Corpora*, Paul Baker & Anthony McEnery (eds), 266–284. Houndmills: Palgrave Macmillan. https://doi.org/10.1057/9781137431738_13
- Karlík, Petr. 2017. Přísllovečné určení času. In *CzechEncy – Nový encyklopedický slovník češtiny*, Petr Karlík, Marek Nekula & Jana Pleskalová (eds). <<https://www.czechency.org/slovník/PŘÍSLOVEČNÉURČENÍČASU>> (23 November 2019).
- Knowles, Murray & Malmkjaer, Kirsten. 1996. *Language and Control in Children's Literature*. London: Routledge.
- Lindquist, Hans & Levin, Magnus. 2008. Foot and mouth. The phrasal patterns of two frequent nouns. In *Phraseology. An Interdisciplinary Perspective*, Sylviane Granger & Fanny Meunier (eds), 143–158. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.139.15lin>
- Mahlberg, Michaela. 2013. *Corpus Stylistics and Dickens's Fiction*. London: Routledge.
<https://doi.org/10.4324/9780203076088>
- Navratil, Jiri & Werner Zuhlke. 1997. Double bigram-decoding in phonotactic language identification. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, 1115–1118. IEEE. <https://doi.org/10.1109/ICASSP.1997.596137>
- Nikolajeva, Maria. 2004. Narrative theory and children's literature. In *International Companion Encyclopedia of Children's Literature*, Vol. 1, 2nd edn, Peter Hunt (ed.), 166–178. London: Routledge.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Pearson Longman.

- Sainsbury, Lisa. 2014. Chronotopes and heritage: Time and memory in contemporary children's literature. In *Modern Children's Literature. An Introduction*, 2nd edn, Catherine Butler & Kimberley Reynolds (eds), 187–201. Houndmills: Palgrave.
https://doi.org/10.1007/978-1-137-36501-9_13
- Schukat-Talamazzini, Ernst-Günter, Hendrych, R., Kompe, Ralf & Niemann, Heinrich. 1995. Permugram language models. In *Fourth European Conference on Speech Communication and Technology (EUROSPEECH)*, 1773–1776. Madrid.
- Šebestová, Denisa & Malá, Markéta. 2019. Expressing time in English and Czech children's literature: A contrastive n-gram based study of typologically distant languages. In *Language Use and Linguistic Structure. Proceedings of the Olomouc Linguistics Colloquium 2018*, Joseph Emonds, Markéta Janebová & Ludmila Veselovská (eds), 469–483. Olomouc: Palacký University.
- Sinclair, John. 2004. *Trust the Text*. London: Routledge. <https://doi.org/10.4324/9780203594070>
- Stubbs, Michael. 2007. An example of frequent English phraseology: Distributions, structures and functions. In *Corpus Linguistics Twenty-Five Years On*, Roberta Facchinetti (ed.), 89–105. Amsterdam: Rodopi.
- Thompson, Paul & Sealey, Alison. 2007. Through children's eyes? Corpus evidence of the features of children's literature. *International Journal of Corpus Linguistics* 12(1): 1–23.
<https://doi.org/10.1075/ijcl.12.1.03tho>
- Wild, Kate, Kilgariff, Adam & Tugwell, David. 2012. The Oxford Children's Corpus: Using a children's corpus in lexicography. *International Journal of Lexicography* 26(2): 190–218.
<https://doi.org/10.1093/ijl/ecso17>

Corpora and Tools

- Křen, Michal et al. 2018. *Korpus SYN-7*. Prague: Ústav Českého národního korpusu FF UK.
 <<http://www.korpus.cz>>
- Milička, Jiří. 2019. *Engrammer*. <<http://milicka.cz/en/engrammer>>
- BNC Consortium. 2007. *The British National Corpus*, XML edition. Oxford Text Archive.
 <<http://hdl.handle.net/20.500.12024/2554>>

Index

A

- adjunct 203, 205, 207–225, 231–233, 245, 249
 see also time adjunct
- adverbial placement 5, 201–202, 204, 213, 246, 248
- argumentational 181, 186–191, 196–198
- aspect 1–2, 4, 7, 95–106, 108–116, 118–124, 156, 172–173, 177, 261–262, 267, 273, 278, 286
- aspectual particle 99, 101, 103–104, 108, 113–114, 116, 118–119, 121
- auxiliarization 4, 86, 88–90

B

- backgrounding 188, 193
- BE06 corpus 205–206, 210–214, 217, 219–220
- binomial 3–5, 39–60, 62–65, 129–130, 132–153
 see also echoic binomial, gendered binomial

C

- children's fiction 6, 44–45, 62, 283–288, 290, 292–293, 298–302
 see also fiction
- cognitive accessibility 44, 60
- coherence 183–184, 198, 203
- colligation 3, 202, 204–205, 216, 218–220, 222–224, 293
- collocation 3–4, 6–7, 9–10, 12–17, 19–21, 33–35, 57, 182, 188, 202, 204–205, 218, 220, 222–223, 250, 289, 295
- comparable corpora 3, 5, 182, 184–185, 202, 206, 286, 295, 302
- concessive 181, 188, 197
- conditional 181, 191–193, 195, 197, 204

- contrastive analysis 6, 69, 156, 184–185, 224, 248, 301
- contrastive study 4, 8, 40, 95, 129, 132, 178, 202–203, 224, 226, 229–230, 232, 234, 245, 247–248, 283
- Czech 4, 6, 39–41, 44–46, 51–52, 56–59, 61–62, 283–290, 292–295, 299–302

D

- deictic time adverbials 202
 see also time adverbial
- disciplinary variation 257–261
- discourse pattern 181, 183–184, 186–187, 190, 197

E

- echoic binomial 4, 129, 133–134, 136, 139, 141, 151–152
 see also binomial
- English-Norwegian Parallel Corpus 67, 93, 155, 229
 see also ENPC
- English-Swedish Parallel Corpus 68, 129–130
 see also ESPC
- Engrammer 6, 283–285, 290–291, 301–302
- ENPC 67–73, 77–79, 81, 90, 155–156, 159–161, 165, 168, 170, 172–173, 175–176
 see also English-Norwegian Parallel Corpus
- ESPC 68, 70, 72, 131, 142–145
 see also English-Swedish Parallel Corpus
- expression of time 2, 101, 203, 285
 see also temporal expression

F

- fiction 6, 44–45, 56, 60, 62, 131, 143–145, 155, 159–161, 170, 202–204, 219, 223,

- 229–230, 232, 283–288, 290, 292–293, 295, 298–302
 see also children's fiction
- final position 6, 196, 204, 229, 231, 233, 238, 245–248, 289
 see also initial position, medial position
- focus 26, 77, 99–100, 187, 192–194, 197–198, 201, 212, 241, 244, 247
- football match report 6, 229–233, 235, 247–248
- foregrounding 189, 192–193, 197
- French 2–5, 61, 95, 97–101, 103, 105–111, 114–123, 181–182, 184–187, 189–192, 194–198, 205, 229
- frequent phrases 222, 235

G

- gendered binomials 42, 46–47, 50, 52, 54, 62
 see also binomial
- German 2–5, 40, 73, 130–142, 144–146, 148–152
- grammaticalization 2, 4–5, 7, 67, 69, 81–83, 86, 88–90, 104, 110, 181, 188, 190, 195, 197

H

- historic present 6, 239, 246, 248

I

- Inference 181, 191, 193, 195, 197
- initial position 6, 187–188, 203–205, 208, 215, 221–222, 229, 231–233, 239, 245–247
 see also medial position, final position
- interval 5, 155–159, 160–162, 165–169, 171, 173, 176–177, 242, 269, 290

K

kinship term 4, 40–41, 43,
46–49, 51, 53–54, 58–60,
62–63

L

L1 59, 256, 261–262, 264–265,
272, 280
L2 256, 260–262, 265, 280
LEGS 130–132, 134–137,
139–144, 148, 150–151
see also Linnaeus University
English-German-Swedish
corpus
lexical priming 5–6, 201–202,
204–205, 218–220,
223–224
lexical realization 5, 130,
201–202, 216–218, 223–224
Linnaeus University English-
German-Swedish
corpus 129–131
see also LEGS
Liverpool 9–11, 13–33, 35

M

Mandarin 3–4, 95–97, 99,
101–123
marked 4, 6, 41–42, 59, 62,
95, 102–103, 106, 108–123,
184, 215, 233, 258, 261–262,
267, 273
see also unmarked
markedness 42, 58, 60, 62
master's theses 6, 255–259,
264–265, 279–280
medial position 5, 186, 201,
204, 209, 213–216,
221–223, 238
see also initial position, final
position
methods section 6, 255–257,
260, 263–265, 267, 269,
278–279
minute 6, 158, 166, 168, 175,
229–248, 288, 298–299
modality 4, 67–69, 78, 87, 91
multilingual corpora 3–4, 7–8,
71, 92, 95, 178, 226
multi-word combination 156,
171–172, 174–176
multi-word unit 3, 6,
283–285, 289

mutual correspondence

169–170, 177

N

negation 89–90
news discourse 5, 201–202
newspaper 9–11, 16, 19, 26,
32–33, 35, 85, 201–202,
205–206, 219, 223, 227,
229, 231–232, 235, 237–238
n-gram 3, 6, 203, 232, 235, 283,
285–296, 298, 300–302
nineteenth century 9, 11, 17,
25, 27–29, 35, 40
noisy data 11, 13, 16
non-fiction 56, 131, 136,
143–145, 155, 160–161, 170
NPN construction 130,
132–135, 139–140, 151–152

O

OCR error 11–13, 16–17, 20

P

Paradigmatization 86
parallel translations 70–71
participant 4, 68–69, 77,
79, 88, 90–91, 107, 231,
240–242, 246–248
past tense 6, 98, 119, 239–240,
246–248, 260–262, 266,
270–272, 274–279
perceptual markedness 42, 62
phraseme 183, 198
phraseology 3, 5–6, 41, 44,
46, 129–132, 146–147, 149,
151–152, 182–184, 202, 205,
216, 219, 230, 232–233, 237,
244, 247, 258, 283–284, 289
point in time 5, 12, 155–156,
158–159, 161–162, 165–166,
244, 246
polysemy 5, 67, 71, 82, 90, 155,
166, 169, 181, 184, 186–187,
189–190, 196, 198, 286
presentational 187,
194–195, 197

R

rhetorical function 134,
183, 192, 196, 260,
262–263, 265–267, 272,
274–276, 299

S

semi-auxiliary 4, 67, 69, 71, 73,
76, 82, 87
skipgram 290
slave trade 9–10, 17–18,
20–35
slavery 4, 9–10, 14, 18, 20,
24–29, 31, 35
space-time semantic
change 83
student academic writing 257,
263, 265
sufficiency 4, 67–69, 71–72,
74–82, 85, 87–91
see also time sufficiency
Swedish 3–5, 40, 67–70, 72, 75,
81–83, 85, 89–90, 129–132,
135–146, 148–152, 174, 204,
207, 213

T

temporal constructions 158,
302
temporal expression 3, 68, 76,
89–90, 156, 158, 176, 184,
187, 192, 202–203, 232,
286–289, 290, 294, 301
see also expression of time
temporal reference 89, 97, 101,
109, 120–121, 297–298
tense 1–4, 6–7, 95–98, 100–103,
105–106, 108–109, 111–116,
118–124, 187, 231, 238–240,
246–248, 255–257,
260–263, 265–266,
268–279, 286
see also verbal tense
tenseless language 4, 95–97,
103, 106, 123
text type 40, 51, 54, 160, 201,
229–231, 232–235,
245–246, 248, 259, 265,
273, 286
time adjunct 207–219, 223–224,
231–232
see also adjunct, time
adverbial
time adverbial 5, 107, 201–206,
208, 230, 244, 289
see also time adjunct
time sufficiency 4, 67–69, 71,
76–82, 85, 87, 90–91
topic shifter 181, 187

- translation corpus 67, 71,
78, 97, 105–107, 119,
123–124
- translation correspondence 3,
71, 74, 79, 108, 130, 132,
148–149, 152, 156, 165,
175–176
- translation universals 105–106,
124, 144, 147
- Translationese 144
- triangulation 10, 17, 34
- U**
- UFA 9–10, 13–17, 19, 21, 34–35
see also Usage Fluctuation
Analysis
- unmarked 41–42, 62
see also marked
- unordered n-grams 6, 290, 301
see also n-grams
- Usage Fluctuation Analysis 4,
9–10
see also UFA
- V**
- verbal tense 4, 95, 97–98, 100,
105–106, 108–109, 111–116,
118–124, 286
see also tense
- voice 261–262, 267,
273–274
- Z**
- zero translation 109, 166, 169,
170, 173, 175, 177

This volume comprises a collection of contrastive studies on language and time. Languages represented include Czech, French, German, Mandarin, Norwegian and Swedish, all of which are contrasted with English. While the amount of published research on temporal relations in general is considerable, less work has been carried out on comparing how we talk about time in various languages and how languages change over time. Several methodological challenges are addressed and solutions proposed, such as how to deal with poor quality historical data and how to identify n-grams in typologically different languages for purposes of comparison. The results of the various studies show how multilingual corpora can increase our knowledge of language-specific features as well as linguistic, typological and cultural differences and similarities across languages.

ISBN 978 90 272 0968 9



9 789027 209689

JOHN BENJAMINS PUBLISHING COMPANY