

**Understanding English-German
Contrasts – A Corpus-based Comparative
Analysis of Ellipses as Cohesive Devices**

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes

vorgelegt von
Katrin Menzel
aus Großröhrsdorf
Saarbrücken, 2016

Erstberichterstatter: Univ.-Prof. Dr. Erich Steiner
Zweitberichterstatter: Univ.-Prof. Dr. Ingo Reich
Dekan: Univ.-Prof. Dr. Roland Marti
Tag der letzten Promotionsleistung: 05.12.2016

Understanding English-German Contrasts -- A Corpus-based Comparative Analysis of Ellipses as Cohesive Devices

Abstract

The objective of this dissertation is to develop a sound conceptual basis for a cross-linguistic study on ellipses as omission phenomena in English and German. The dissertation has a particular focus on endophoric ellipses as cohesive devices creating textual links and contributing to textual coherence. It takes a relatively theory-neutral, broadly functional approach to describing ellipses underpinned with observation statements and examples of real-life ellipsis use based on data from a bilingual corpus of written and spoken language. In this way, the present study has the potential to arbitrate between competing academic theories that concern the nature of ellipses. The choice of focus in the dissertation is motivated by the fact that cohesive devices, including ellipses, are an important topic in text-linguistics and discourse analysis. However, one could rightfully feel dissatisfied with the current lack of a coherent overarching taxonomy for ellipses as cohesive devices. Previous work on ellipses as a broad class of reduced and non-canonical syntactic constructions has often been limited by inconsistent descriptions, conflicting theories or the micro-level analysis of particular isolated phrases or sentences with little practical relevance. It can be observed that ellipses in many cases are not very well defined in the available literature on discourse analysis and textuality where we frequently find rather vague or sketchy explanations of ellipses or examples that, in my opinion, do not always illustrate the concept of

ellipsis correctly. These problems of definition represent a research gap which hinders the development of clear hypotheses and the possibility of drawing meaningful conclusions from the analysis of data. This study provides a long-needed systematic reassessment of the ellipsis description suggested by Halliday and Hasan in the context of a growing and diverse body of literature on elliptical structures.

In this study, I analyse data from a corpus as the more recent studies in contrastive linguistics have emphasised the importance of corpus-based methodologies. Authentic corpus examples embedded in their respective textual contexts are a better way to describe a given language pair than the discussion of simplified, theoretical examples that may restrict our understanding of grammar based on idealised language use and personal intuition. Corpus data in contrastive linguistics allow us to study the frequencies of certain patterns in English compared to German or in certain text types. Comparative corpora and aligned parallel corpora expose learners to authentic language contexts in a structured way and are indispensable for the development of realistic and learner-adequate language models. In order to understand ellipses as cohesive devices, it is necessary to extend the analysis beyond an enumeration of simplified, theoretical examples that rarely occur in actual data, as if language was an abstract entity, not embedded in rich discourse contexts and actual contexts of language use. This study emphasises the value of corpus studies as the patterns and distribution of elliptical structures found in the data partly deviate from assumptions and standard examples from the theoretical literature. Therefore, in this dissertation the frequencies, types and functions of ellipses in an English-German corpus of

written and spoken language are analysed in order to lay the foundations for a discourse-oriented contrastive grammar on textual cohesion and coherence with relevance to theoretical and applied linguistics, translation studies and foreign language pedagogy.

Acknowledgements

As this dissertation project developed over more than five years, I had to rely on support and advice from many people without whom I could not have completed it. I owe the greatest debt of gratitude to my supervisor Professor Dr Erich Steiner from the Department of Applied Linguistics, Translation and Interpreting at Saarland University who has been a true mentor and a constant source of support. He encouraged me to select a topic I felt passionate about and in which I would not lose interest. I have benefited greatly from his expertise in areas such as English syntax, contrastive linguistics, cross-linguistic corpus studies and the properties of translated texts as well as from his profound knowledge of functional approaches to discourse analysis, textual cohesion and register theory. He was always available to give me thoughtful feedback which was extremely valuable and helpful. I was given the opportunity to present the progress of my work and the results of my research at numerous symposia, project meetings and scientific conferences and under his guidance I was able to publish my first journal articles and conference papers on subjects that are related to specific aspects of my research.

I have also benefited greatly from the comments and expert guidance on the topic of ellipsis by my second advisor Professor Dr Ingo Reich from the Department of German Language and Literature at Saarland University who added substantially to my understanding of the issues addressed. I am grateful for the various discussions and the invitation to a PhD colloquium at the Department of German Language and Literature where I had the opportunity to present my

thesis project and to receive relevant feedback on the subject of ellipsis. I particularly thank Eva Horch for sharing her time and insights.

This dissertation contributes to the DFG-funded Saarbrücken “GECCo” project (“Kohäsion im Deutschen und Englischen – ein empirischer Ansatz zum kontrastiven Vergleich”) and I gratefully acknowledge the financial support by the German Research Foundation and Saarland University that enabled me to present my work at national and international conferences, for instance at a panel on innovation in discourse analytic approaches to translation studies by Jeremy Munday and Meifang Zhang at IATIS in Belo Horizonte. Taking part in such conferences allowed me to expand my scientific network and to receive valuable feedback. This funding allowed me to take part in various workshops such as the “Parenthesis and Ellipsis” workshop by Marlies Kluck, Dennis Ott and Mark de Vries at the Annual Conference of the German Linguistic Society (DGfS) in Potsdam and the pre-conference workshop by Vaclav Brezina on spoken corpora at the Corpus Linguistics conference (CL2013) in Lancaster. My heartfelt thanks go to the current and the former members of our GECCo project who have all put much hard work and dedication into this research project. I thank Erich Steiner, Kerstin Kunz, Ekaterina Lapshinova-Koltunski, José Manuel Martínez Martínez, Stefania Degaetano-Ortlieb, Marilisa Amoia and our student assistants for their contributions to various aspects of this work, for their support, and for creating a cordial working environment.

Since 2011, I have belonged to the Department of Applied Linguistics, Translation and Interpreting (now: Department of Language Science and Technology) at Saarland University as a research assistant and lecturer. I still

think fondly of my time as an undergraduate student in the same department when I was a student assistant working on GECCo's predecessor project CroCo. I sincerely thank Stella Neumann, Silvia Hansen-Schirra and Kerstin Kunz who awakened my interest in corpus-based research and contrastive linguistics and whose academic accomplishments continually inspire me. I thank Bert Vaux at the Department of Theoretical and Applied Linguistics of Cambridge University where I spent a year as an undergraduate student for his well-illustrated and entertaining lectures, which made his students realise that linguistics is FUN!

I also wish to thank a number of academics who gave their valuable comments on my work at PhD colloquia and in our discussions, particularly Elke Teich and Augustin Speyer. I have also greatly benefited from the interest and encouragement that I have received from a wide range of other researchers and colleagues including Marco Agnetta, Mirja Hanke, Ashraf Khamis, Hannah Kermes, Jörg Knappen, Marie-Ann Kühne, Karin Maksymski, Daniele Moretti, Santanu Pal, Heike Przybyl, Anne-Kathrin Schumann, Liling Tan, Mihaela Vela, Anne Weber, Andrea Wurm and Marcos Zampieri. I also thank Marc Summkeller who rescued me from many computer-related pitfalls and promptly solved any technical issues, and I thank Peggy Daut and Ulrike Konz for their constant encouragement and unbelievable support. I am also grateful to my talented students, several of whom wrote their term and graduation papers on various aspects of textual cohesion and who, no doubt, have assisted my development with their lively and insightful discussions.

I am thankful for Zell's SPSS seminar, the GradUS-financed course on MATLAB and Markus Pospeschill's consultation on statistical concepts and

analyses as well as for the Clarin-D PhD student days on corpus-linguistic methods. I received Erasmus travel grants to visit universities in Izmir, Joensuu, Edinburgh and Iași. I would like to extend my thanks to the colleagues that I met during these research and teaching visits, especially Hülya Kaya, Leena Kolehmainen and Marion Winters. Various other friends and colleagues have inspired and encouraged me. My very special thanks go to my dear friend Dorothy Brito from the Universidade Federal Rural de Pernambuco, an excellent expert on syntax, and Ben Clarke from the University of Portsmouth who enabled me to take advantage of his expert knowledge on ellipsis in the systemic functional literature. I am particularly grateful to Kathy Saranpa for proofreading the manuscript. Any remaining errors are, of course, my own.

I would like to thank Olivier Landeville who endured endless stories of my research progress and finally, I would like to thank my family, especially my parents and my brother. Words cannot express the appreciation I have for their love and support.

Table of Contents

List of Tables	xii
List of Figures	xiii
1. Introduction	1
1.1 Motivation, background and research questions	1
1.2 Overview of chapters.....	11
1.3 The GECCo project.....	17
1.4 Assumptions & hypotheses on frequencies of ellipses in English / German	20
2. A brief historiography of ellipsis as a grammatical concept	32
3. The description and analysis of elliptical phenomena	41
3.1 Ellipsis and the written-spoken dimension	41
3.2 Ellipsis and the syntax-semantics-pragmatics interface	46
3.3 Ellipsis scope, non-identity effects and acceptability	57
4. The classifications of ellipsis subcategories	74
4.1 Existing English and German ellipsis taxonomies	74
4.2 Ellipsis as a cohesive device in the systemic functional approach	91
5. Ellipses as cohesive devices	110
5.1 Nominal ellipses	110
5.1.1 The noun phrase in English and German	110
5.1.2 Cohesive and non-cohesive cases of nominal ellipses	128
5.1.3 Nominal ellipses after numerals	141
5.1.4 Nominal ellipses after adjectives.....	146
5.1.5 Nominal ellipses after possessives and after classifier nouns	153
5.1.6 Nominal ellipses after quantifiers	157
5.1.7 Nominal ellipses vs. nominalisation	161
5.1.8 Nominal ellipses vs. Right Node Raising and ‘ellipsis’ within the NP.....	170
5.2 Verbal ellipses.....	177
5.2.1 The verb phrase in English and German.....	177
5.2.2 Cohesive and non-cohesive cases of verbal ellipses	189
5.2.3 Operator ellipses.....	204
5.2.4 Lexical verb ellipses	209
5.2.5 Verbal ellipses vs. verbal substitution	228
5.2.6 The distinction between verbal and clausal ellipses.....	236
5.3 Clausal ellipses.....	239
5.3.1 The clause in English and German	239
5.3.2 Adjacency ellipses	243
5.3.3 Sluicing	249
5.3.4 Omissions of predicative expressions.....	253
5.3.5 Mixed cases.....	256
5.3.6 The difference between clausal substitution and clausal ellipsis	258
6. Other fragments and reductions	263
6.1 Overview.....	263
6.2 Sentence splits.....	266
6.3 Answering particles.....	275

6.4 Non-clausal units.....	280
6.5 Text-type-specific fragments	284
6.6 Other	290
7. The relationship between ellipsis and other cohesive devices	296
8. Corpus resources.....	306
8.1 Previous corpus-based studies on ellipses	306
8.2 GECCo – details on corpus design and compilation.....	313
8.3 Annotation, retrieval and extraction of ellipses	329
8.4 Manual vs. automatic annotation procedures for cohesive ellipses	339
9. Analysis of corpus data and results.....	349
9.1 Overview, extracted results and computational tools.....	349
9.2 Cohesive ellipses in English and German	355
9.3 Differences between originals and translations	374
9.4 A comparison of written and spoken mode	388
9.5 Register variation	394
10. Conclusions and outlook	409
11. Zusammenfassung der Dissertation (Summary in German).....	413
References	424
Annex 1: Frequencies of ellipses and fragments in GECCo per text	452

List of Tables

Table 1: Ellipsis subtypes according to different authors.....	84
Table 2: Size of GECCo subcorpora	318
Table 3: Size of GECCo registers	319
Table 4: Query examples used to extract the categories under analysis.....	338
Table 5: Frequencies of all cohesive ellipses, cohesive nominal ellipses and cohesive verbal/clausal ellipses in English & German originals	355
Table 6: Frequencies of all non-cohesive ellipse, non-cohesive nominal ellipses, non-cohesive verbal/clausal ellipses and other fragments in English & German originals.....	368
Table 7: Frequencies of all cohesive ellipses, cohesive nominal ellipses and cohesive verbal/clausal ellipses in originals (written texts) and translations in GECCo.	374
Table 8 : Frequencies of all cohesive ellipses, cohesive nominal ellipses and cohesive verbal/clausal ellipses in written and spoken data in GECCo	389
Table 9: Frequencies of nominal and verbal/clausal cohesive ellipses per register	395
Table 10: Frequencies of non-cohesive ellipses and fragments per register.....	399
Table 11: Number mismatches between nominal ellipsis and nominal antecedent	405

List of Figures

Figure 1: GECCo corpus architecture	19
Figure 2: The structure of the noun phrase according to Hill (1958)	113
Figure 3: Zones of premodification according to Quirk et al. (1985)	114
Figure 4: The structure of the English noun phrase according to Feist (2011)	115
Figure 5: The structure of the nominal group according to Matthiessen et al. (2010) ..	115
Figure 6: System network of the clause in English (Matthiessen and Teruya, 2010) ..	239
Figure 7: Size of GECCo subcorpora	320
Figure 8: Size of GECCo registers	320
Figure 9: Number of tokens per text in GECCo subcorpora	322
Figure 10: Annotation categories of potentially cohesive ellipsis in MMAX2	330
Figure 11: Annotation of function of ellipsis in MMAX2	330
Figure 12: Link between elliptical phrase and antecedent in MMAX2	332
Figure 13: Annotation of elements as ellipses or antecedents	332
Figure 14: Nominal ellipsis with anaphoric reference to antecedent occurring both in the same clause and in one of the previous sentences	334
Figure 15: MMAX2 annotation of an example of a non-clausal unit	335
Figure 16: Number of cohesive ellipses per text in EO and GO	356
Figure 17: Scatter plots for text length vs. absolute number of coh. ellipses EO/GO ..	359
Figure 18: Cohesive nominal ellipses in EO and GO	361
Figure 19: Cohesive verbal/clausal ellipses in EO and GO	363
Figure 20: Cohesive nominal and verbal/clausal ellipses in texts in EO/GO	365
Figure 21: Cohesive nominal and verbal/clausal ellipses in EO/GO	365
Figure 22: Scatter plots for cohesive nominal and verbal/clausal ellipses in EO/GO ..	366
Figure 23: Cohesive ellipses, non-cohesive ellipses and other fragments in EO/GO ..	369
Figure 24: Scatter plots for cohesive and non-cohesive ellipses in EO and GO	370
Figure 25: Notched boxplot of fragments in texts in EO and GO	371
Figure 26: Scatter plots for cohesive ellipses and fragments in EO and GO	372
Figure 27: Frequencies of cohesive ellipses in original texts and translated texts	375
Figure 28: Cohesive ellipses in original and translated texts	376
Figure 29: Frequencies of cohesive ellipses in pairs of original and translated texts ..	377
Figure 30: Percentage of cohesive ellipses from original texts that correspond to similar syntactic structures in sentence-aligned translations	385
Figure 31: Frequencies of cohesive nominal ellipses, cohesive verbal/clausal ellipses and sum of these types in written and spoken subcorpora	390
Figure 32: Frequencies of cohesive ellipses in spoken corpus sections	392
Figure 33: Proportions of cohesive ellipses in English and German data	396
Figure 34: Proportions of nominal cohesive ellipses in English and German data	397
Figure 35: Proportions of cohesive verbal/clausal ellipses in English & German data	397
Figure 36: Proportions of fragments in English and German data	401
Figure 37: Distributions of different cohesive devices per language and register	403

1. Introduction

1.1 Motivation, background and research questions

“Ellipsis continues to fascinate because its analysis goes directly to the heart of the main reason we study syntax: to discern the nature of the form/meaning correspondence.”

(Merchant, 2016)

Ellipsis is an important topic in syntactic research. It is also a topic closely tied to formal and semantic patterns at the discourse level. This dissertation has its focus on English and German ellipsis-antecedent relations that contribute to the cohesiveness of texts. Cohesion, one of the features of textuality, is achieved by the use of cohesive devices. These devices are mainly elements that appear in the surface structure of texts as indicators of lexico-grammatical relations across clause or sentence boundaries. In de Beaugrande and Dressler (1981: 3) and in Baker (1992: 180), cohesion is defined as a textual surface relation, a network of lexical and grammatical links between various parts of a text.¹ In the case of ellipsis, one of the types of cohesive devices suggested by Halliday and Hasan (1976), a textual relation is not exactly set up between two actual text segments, but between a textual antecedent and a subsequent omission

¹ Textual cohesion in multilingual contexts is discussed in Menzel et al. (forthcoming, 1-11).

(cf. de Beaugrande, 1991: 252). In contrast to other cohesive ties, ellipses as cohesive devices involve omitted elements that are assumed to be present in the underlying syntactic structure. Normally, when categories are equated across languages, surface forms are compared. With ellipsis, we cannot compare surface forms of overt structures. We have to compare the structural patterns of the contexts in which omissions occur and we have to derive the internal structure of ellipses from abstract, underlying structures. In this study, I am interested in omissions which are something invisible and whose full underlying structures can only be deduced from the surrounding contexts, but it is a challenge to describe the 'ghost structures' of the unspoken. My approach is based on the assumption that languages have 'deep' or underlying structures, detectable only by inference. Some phrases and sentences are remnants of full constructions, of which certain parts have been annihilated by ellipses. An ellipsis creates a space of possibility, an invisible, potential form of syntactic 'reality' that one can infer and identify from the form of its remnant structure.

One of the key objectives of contrastive linguistics is to raise awareness of contrasts and similarities between different languages, carrying implications for theoretical linguistics, foreign language pedagogy, translator training and multilingual text production. In order to understand how a language works, including the structures and uniqueness of one's own language, it is worth reflecting on how things could be different or how they are organised in other languages. With regard to the English-German language pair, several overviews of general linguistic contrasts

exist. Such overviews often emphasise the close historical or ‘genetic’ relationship between English and German as descendants of the same branch of Germanic that explains many similarities between these two languages. At the same time, they point out the historical rift between these languages as well as more or less intensive contact with other languages and processes of language-internal innovation that may account for various differences between English and German. Existing references in English-German contrastive linguistics have been listed in Markus and Wallmannsberger (1987), in König (2001) and in Hall (2010). One of the earliest comprehensive works on grammatical contrasts between English and German was Kufner (1962). Among the most widely cited and most frequently used books on English-German contrasts are Hawkins’ *Comparative Typology of English and German* (1986, cf. also Hawkins, 1988) and König and Gast’s textbook (2012). In both works, various contrasts on different linguistic levels, e.g. phonological, morphological, syntactic and semantic aspects, are discussed. Beck and Gergel (2014) address a selection of semantic and syntactic issues in their monograph on contrasts between English and German. Königs (2011) presents a practice-oriented description of a selection of English grammatical structures that may pose difficulties for German translation students. These overviews have become useful reference works in English and German linguistics and in translator training. Among the aspects that have been focused on in such works with mainly pedagogical and practical dimensions are the position of the verb in the sentence and the consequences this implies as

well as word-order freedom, participle constructions, infinite constructions or cleft sentences. Some of these references have a stronger theoretical character and address general theoretical questions, for instance, whether the German morphological and syntactic surface structures are in a closer correspondence with their associated meanings and whether there is greater ambiguity in the mapping between surface forms and their meanings in English due to the reduced morphology in English (Hawkins, 1986: 121). Hawkins assumes that in each area of grammar, either the German structures include the English ones or vice versa. Another question that has been raised is whether English grammar is generally more explicit than German grammar (Kortmann and Meyer, 1992). Certain cross-linguistic studies focus on specific areas of contrast such as German-English verb valency or sentence structures (Fischer, 1997, 2013), German focus particles and their English equivalents (König, 1982) or the pragmatic features of certain politeness markers in both languages (e.g. House, 1989). A closer look at the existing literature on English-German contrasts reveals that the similarities and differences between English and German which have been described in the past are predominantly concerned with isolated or sentence-internal grammatical phenomena, neither taking discourse organisation and textual cohesion nor register-specific differences sufficiently into account. The existing general overviews on cross-linguistic aspects illustrate a wealth of differences at various linguistic levels and sometimes do not discuss the individual areas of cross-linguistic differences in great depth and granularity. More specific

contrastive overviews, on the other hand, that focus on particular grammatical topics often describe only a selection of aspects in the length of a journal article, but rarely in monograph-length studies.

There is, as yet, no cross-linguistic overview of ellipses in English and German, and little attention has been paid so far to the discourse function of ellipses and to the role they play as text-building devices in English and German. To fill this existing research gap, this dissertation aims to set out a conceptual framework for the cross-linguistic analysis of cohesive ellipsis. Ellipses as cohesive devices creating textual links have been studied less extensively than other cohesive phenomena both in monolingual and in cross-linguistic studies, which is probably due to their conceptual complexity and their high degree of variation despite relatively low textual frequencies. Large amounts of textual data are necessary to study frequencies and distributions of ellipses as a rather rare phenomenon and a specific stylistic device.

The present corpus-based study aims at describing similarities and contrasts in the use of cohesive ellipses dependent on:

- language (English vs. German)
- text production type (original texts vs. translations)
- mode (written vs. spoken language) and
- different registers / text types.

In the context of this cross-linguistic study, the remnants of endophoric

ellipses in connection with their textual antecedents have been annotated in the English-German electronic text corpus “GECCo”. I annotated ellipses as cohesive devices along with other types of non-cohesive ellipses and sentence fragments on the basis of the annotation scheme that I describe in this dissertation. Due to the widely varying structures of ellipsis remnants, manual annotation turned out to be more efficient and reliable than the semi-automatic procedures that I tested. After the annotation I performed on the corpus data, the annotated patterns have been extracted from the corpus and evaluated with statistical methods to analyse the differences between languages, text production types, modes and registers. My main research questions are:

- How frequent are cohesive ellipses and their subtypes in the dataset under analysis?
- Where do differences and similarities lie: between languages, between registers, between spoken and written modes or between originals and translations?
- Where do we observe the biggest differences?

The strong empirical focus in contemporary linguistics has drawn the attention of researchers to large textual corpora and to quantitative methods as a means of analysis of discourse phenomena. The challenge

with many linguistic terms such as ellipsis is that they do not easily provide a basis for empirical research as they have evolved from philosophical and rhetorical concepts and still lack uniform and exact definitions. They have been and are still used as semi-popular terms with vague boundaries, and their usage is often inconsistent. Existing typological schemes do not always place ellipsis subtypes on an equal footing and often turn out to have grey areas or overlaps between their categories. This makes them highly non-operational for corpus linguistic analyses. Throughout the research for this study, it became increasingly clear that a conceptual clarification and an updated classification system which is more nuanced than previously suggested relatively vague descriptions of ellipses as cohesive devices are necessary before turning to an empirical analysis on cross-linguistic similarities and differences in the use of textual cohesion. The current lack of a coherent, overarching taxonomy for cohesive ellipses hinders the development of clear hypotheses and the possibility of drawing meaningful conclusions from the analysis of data. This dissertation suggests an enriched and improved taxonomy of ellipses as cohesive ties which is supplemented with insights from a large body of literature and underpinned by a systematic cross-linguistic investigation. To understand ellipses as cohesive devices, it is necessary to extend the analysis beyond simplified, theoretical examples that do not occur in actual data in this form. I include numerous ellipsis examples in the description that are embedded in rich discourse contexts and actual contexts of language use.

A systematic review of the existing literature on ellipsis is part of the research process and methodology in this study in order to clarify some of the ongoing controversies. I will also describe the origin of the concept of ellipsis and trace certain stages in its historical development. The historiography of linguistic concepts is not unimportant or negligible, as it helps to explain how certain definitions have evolved that are now applied in contemporary linguistic studies and put forward in pedagogical grammars. There can be no doubt that research methods such as quantitative analyses have to be based on sound theoretical considerations and on a thorough conceptual analysis. Therefore, this thesis aims to identify deficiencies in current knowledge on textual cohesion, to expose conceptual problems in existing models such as vague or illogical classifications and explanations and to evaluate the consistency of existing accounts.

In this empirical study on ellipses as cohesive devices, we cannot use a clear existing definition from the literature and have to specify and define the concept in operationalisable terms. In an attempt to identify and clarify areas that can lead to misconceptions or confusion, I partly revise and elaborate on Halliday and Hasan's work that is often cited to refer to ellipses as cohesive devices. In its current form, Halliday and Hasan's description of ellipses in a systemic functional framework (1976) has some limitations with regard to its applicability to real discourse data from different text types in English and German and not only to idealised or prototypical isolated utterance pairs. Depending primarily on a basic

linguistic reference unit, ellipses in previous descriptions such as Halliday and Hasan's seem like a relative concept or a grammatical phenomenon on a prototype continuum with core and peripheral members. The classifications used for the corpus annotation should not overlap or involve gradual categories or grey areas. The aim is to place all cases found in the corpus clearly in only one category in order to provide the basis for a meaningful quantitative analysis. I take the tripartite ellipsis typology of nominal, verbal and clausal ellipsis, originally suggested by Halliday and Hasan (1976), as a starting point for the development of a fine-grained taxonomy and annotation scheme for ellipses in English and German. The use of such relatively broad, overarching categories of ellipses such as those suggested by Halliday and Hasan permits broad generalisations to be drawn and provides a particularly suitable framework for a cross-linguistic analysis of ellipses with possible textual antecedents in English and German. The three main categories of ellipsis are broad or abstract enough to instantiate cross-linguistic categories for English and German and to serve as a *tertium comparationis*. This ellipsis taxonomy has traditionally been applied in systemic functional approaches in the context of discourse analysis and studies on textual cohesion with regard to various languages, but the concepts have often been used in a slightly superficial and relatively unquestioned manner involving conceptual inconsistencies. In other approaches that discuss ellipses, Halliday and Hasan's description of cohesive ellipses has received little attention so far. Part of the reasons for this lies in the fact that a precise, detailed conceptual understanding of

ellipses as cohesive devices has not yet been fully developed and Halliday and Hasan's ellipsis model has not been elaborated on to any significant extent or tested sufficiently with actual data to show whether it can serve as a viable model. This may explain why theoretical work on the conceptualisation of ellipses in general has mainly been carried out in other theoretical frameworks within the last decades. What studies on ellipses in non-SFL accounts are often most interested in are the assumed mechanisms of recoverability and interpretation of the ellipsis site (cf. Chapters 3.2 and 3.3) – with no particular focus on the potential function of ellipses to create textual links across clause and sentence boundaries.

On a general level, the aim of this study is to suggest a model that makes it possible to specify and define the traditional ellipsis concept in operationalisable terms for empirical quantitative studies. I would like to lay the foundations for a discourse-oriented contrastive grammar on the English-German language pair with relevance to theoretical and applied linguistics, translation studies and foreign language pedagogy. Translators and language learners need to understand how to use the different types of ellipses appropriately as textual elements and need to gain cross-linguistic awareness of these syntactic patterns as stylistically marked constructions and cohesive devices used with typical frequencies in different communication scenarios.

1.2 Overview of chapters

The discussion in this thesis is organised as follows. After this introductory chapter, which has introduced the context for this study and will provide a brief overview of the corpus resources as well as a detailed discussion of my assumptions and hypotheses, Chapter 2 continues by sketching the history of ellipsis as a grammatical concept from classical antiquity to modern times. To a certain extent, the concepts and categories we now use to analyse and describe modern languages have been shaped by the heritage of early philological, rhetorical and even philosophical discourses. Chapter 2 demonstrates that it is worth looking at how the term ‘ellipsis’ and the related concept of the grammatically complete and correct sentence have evolved from classical antiquity to modern times as a useful step that will help to untangle some of the complex issues regarding this term.

Chapter 3 focuses on changing linguistic theories and methodological prerequisites in the Anglophone and Germanophone academic discourses from the more recent past. It discusses methodological and analytical, qualitative approaches that have been covered in the last decades with regard to ellipsis and addresses several central and general questions related to the research on ellipses. Among the topics that continue to be foci of interest with regard to ellipses are grammatical descriptions of ellipses and irregular sentence types in discussions about the grammar of written vs. spoken language. Identifying discourse units in the latter is particularly difficult and it has been discussed whether the sentence as the

primary unit of language is adequate for capturing the whole picture of spoken language. Additionally, Chapter 3 draws attention to the syntax-semantics-pragmatics interface discussion in which the topic of ellipsis plays a major role. In the literature, different answers have been suggested to the question of whether ellipsis resolution is mainly syntactic, semantic or pragmatic in nature and whether ellipsis indeed involves deletion or not. I address the nature of identity between elided material and textual antecedents and some interactions of scope and ellipsis that pose certain difficulties for syntactic theories.

Chapter 4 is concerned with the different possibilities of establishing ellipsis subcategories and gives an overview on some prominent contemporary ellipsis taxonomies. In this chapter, I discuss challenges for empirical, corpus-based studies arising from existing ellipsis descriptions. In many existing typologies, ellipsis categories tend to overlap or are treated as gradual notions. The myriad of elliptical and fragmentary constructions that have been studied under the notion of ‘ellipsis’ in the past usually do not have a focus on textual cohesion. It is necessary to subsume the multitude of categories suggested in the literature under more general and abstract categories. These categories have to cover the variety of different omission possibilities with the potential to be used as cohesive devices in English and German corpus texts from a wide spectrum of communication scenarios on written-spoken and formal-informal continua. Chapter 4 also provides an overview on the concept of ellipsis as a cohesive device in the systemic functional approach and points out areas

that are still in need of clarification.

Chapter 5 explains the annotation scheme I used for the corpus annotation. It gives a detailed explanation of the cases that fall under nominal, verbal and clausal ellipsis in English and German and explains what distinguishes cohesive ellipses from non-cohesive ellipses and other phenomena such as substitution or Right Node Raising. The annotation scheme is rather detailed and gives numerous examples to cover a broad subset of the range of linguistic structures employed in naturally occurring texts. The purpose of giving many examples is to provide a useful description that can be applied to real discourse data from the corpora and not only to idealised or prototypical isolated utterance pairs that we often find in the literature when ellipsis definitions are discussed. Chapter 5 also provides insights into the process of discussing and deciding on controversial cases in order to develop a consistent annotation standard. The fine-grained ellipsis annotation guidelines that have been developed aim to ensure that all cases found in the corpus can be placed clearly in only one category in order to provide the basis for a meaningful quantitative analysis. The conceptualisation of ellipsis in this study combines several aspects of existing frameworks yet clarifies or adds others. It is based on theoretical research and considerations on ellipses in English and German as well as on the occurrences of ellipses in an extensive and varied data base of our corpus texts from a broad range of spoken and written registers.

Chapter 6 points out that certain types of fragments and non-sentential

units do not necessarily involve an omission and should not be confused with ellipses. Moreover, not all reduction strategies and types of incomplete structures have the potential to be used as a cohesive device. This chapter describes certain types of fragments and reduction strategies as syntactic phenomena that have sometimes been called ‘ellipsis’ in the literature. I view these structures as being conceptually different. They have been annotated for comparative purposes and to clearly distinguish them from cohesive and non-cohesive ellipses.

In Chapter 7, the relationship between ellipsis and other cohesive devices is discussed. Using an endophoric ellipsis is never the only possibility to achieve cohesion in a text. The repetition of lexical material or the use of pro-forms functioning as substitutes are typical ways to avoid elliptical structures. The decision to use an ellipsis, substitution or lexical cohesion to create textual cohesion influences the structure and density of lexical chains, i.e. lexical sequences of semantically related words in a text. In this chapter, I compare the different functions of ellipses, lexical cohesion and substitution.

Chapter 8 summarises previous corpus-based studies on ellipses that are based on different theoretical approaches which have to date been rather small-scale. If elliptical structures were studied on the basis of large existing corpora, the focus was usually not on quantitative results as precision and recall were rather low and the main interest was supporting linguistic statements with a selection of authentic examples. Additionally, this chapter describes the corpus I used and its particular characteristics

and explains the tools and methods related to its compilation as well as ellipsis annotation, retrieval and extraction processes. I compare automatic or manual approaches for the annotation and clarify why I finally opted for a completely manual annotation for ellipses and fragments.

Chapter 9 presents the quantitative results of the corpus study. I first describe and interpret the findings from a comparison of the English and German non-translated data. Then I give an overview on the comparison of original texts and translated texts. Finally, a closer look is taken at the different registers and the differences between written and spoken mode. I provide frequency data and generate summary statistics, visualise these graphically with different types of plots, interpret the data and test various comparative and relationship hypotheses. As expected, the overall frequencies of cohesive ellipses were relatively low in the data in both English and German compared to other types of cohesive devices. Surprisingly, the frequencies of cohesive ellipses in English and German non-translated texts turned out to be almost identical in each corpus section of non-translated texts. Furthermore, the data suggest that the German data have a higher variability than the English data and that ellipses in their function as cohesive devices are generally more frequent in spoken texts in both English and German. Finally, I present my conclusions. The results have several practical implications for language learners and professional writers such as journalists or translators. English and German seem to behave rather similarly when it comes to the use of ellipses in their function as cohesive devices. Nevertheless, it is necessary

to be aware of the functions and differences between ellipses as omissions within and across sentence boundaries and other types of fragments or independent non-clausal units to be able to use these structures appropriately and in language- and register-typical frequencies.

1.3 The GECCo project

The conceptualisation, annotation and analysis of ellipses as cohesive devices contribute to the DFG-funded project “German-English contrasts in cohesion – Towards an empirically-based comparison” (GECCo).² The GECCo project is a comprehensive corpus-based investigation in the field of contrastive linguistics with the aim to identify German-English contrasts and similarities in their systems of cohesion and their textual instantiations of these systems across languages (English vs. German), across registers (different text types and communication scenarios along the written-spoken continuum) and across production types (original vs. translated texts).³ This section gives a brief summary of the GECCo project and a short overview of the GECCo corpus. Although a more detailed description of the corpus will be given in Chapter 8, it is useful to provide some information on the context of the research project at the beginning of this dissertation. This work represents a continuation of a series of studies aiming to identify English-German contrasts in cohesion, and most examples in the following chapters illustrating the theoretical concepts are taken from the GECCo corpus. Some examples are also taken from theoretical discussions about ellipsis or from other sample

² <http://www.gecco.uni-saarland.de/GECCo/index.html>, the GECCo project is funded by the German Research Foundation (DFG) under GZ STE 840/6-1 and 6-2 and KU 3129/1-2.

³ cf. Project Report Work Package WP 3.1.1. (1st project phase, unpublished project-internal report): “We are aiming at frequencies of occurrence depending on language, register, and mode of text production (translation vs. original).”

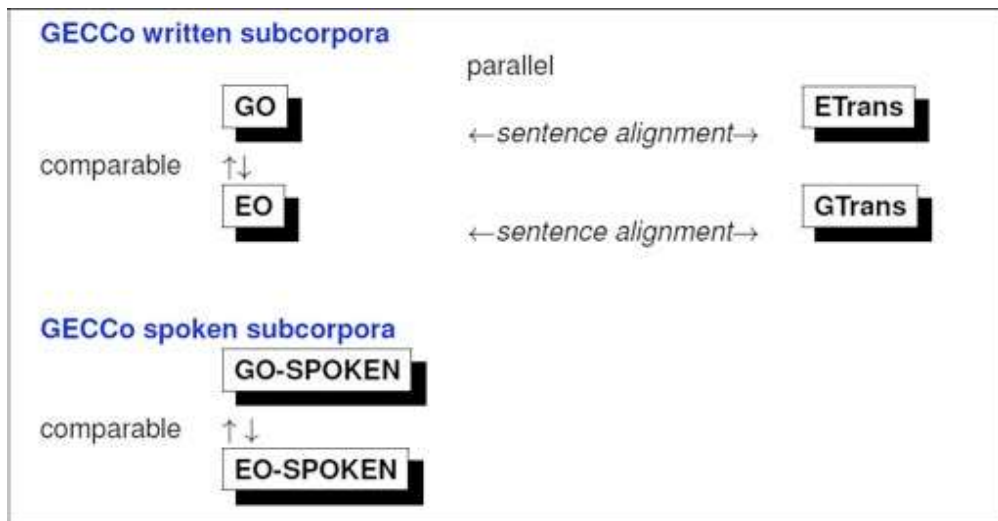
The comparison of original texts and translations plays no major role in the current second project phase that started in 2013, but we will also address this dimension in this dissertation as it was one of the project’s goals in the first project phase when the groundwork for this dissertation was laid.

texts.⁴

The bilingual multilevel-annotated GECCo corpus (ca. 1.69 m. tokens) consists of texts and transcriptions from a broad range of written and spoken registers and text types (cf. Figure 1). GECCo is both a parallel and a comparable corpus. Bilingual *parallel* corpora consist of texts in a language and their translations in another language. Bilingual *comparable* corpora consist of original texts in two languages that are similar with regard to their sampling frame, balance and representativeness (McEnery, 2003: 450). The written part of the corpus is a comparable corpus and at the same time a bi-directional parallel corpus. It contains comparable source texts in English and German as well as their sentence-aligned German and English translations. It is comprised of fictional texts (FICTION), political essays (ESSAY), instruction manuals (INSTR), popular science texts (POPSCI), letters to shareholders (SHARE), prepared speeches (SPEECH), tourism leaflets (TOU) and corporate websites (WEB). The spoken component is a comparable corpus. In an earlier stage of the project, it consisted of two registers, namely academic lectures (ACADEMIC) and interviews (INTERVIEW). The current corpus version includes more spoken registers than before. The recently compiled registers are talk shows (TALKSHOW), internet forums (FORUM), medical consultation (MEDCONSULT) and sermons (SERMON). They

⁴ If example sentences illustrating certain grammatical or ellipsis phenomena were found in books such as novels or non-fiction work and they can be accessed via Google Books, they will be cited directly in the text or in a footnote with reference to the title and the author of the publication and a web link referring to the exact page where this example can be found on Google Books. As these references do not directly contribute to the theoretical discussion of ellipsis, we have decided not to list such references with separate entries in the bibliography section of this work.

have been described in more detail in Kunz et al. (submitted). File headers contain metadata, for instance, data on the speakers, authors and translators and information from a register analysis that has been carried out for each text.









written (imported from CroCo*), texts from 1992 – 2006	
EO 	FICTION, ESSAY
GO 	INSTR, POPSCI
ETRANS 	SHARE, SPEECH
GTRANS 	TOU, WEB
spoken (collected**), texts from 2008 – 2012	
EO 	INTERVIEW, ACADEMIC, FORUM
GO 	TALKSHOW, MEDCONSULT, SERMON

Figure 1: GECCo corpus architecture (corpus version 2016), EO = subcorpus with English originals, GO = German originals, ETRANS = English translations, GTRANS = German translations

1.4 Assumptions and hypotheses on frequencies of ellipses in English and German

This subchapter will summarise the main assumptions and hypotheses related to the frequencies and distributions of ellipses in our English and German data as well as in the different registers, language modes and production types. As already mentioned above, the present study aims at describing similarities and contrasts in the use of cohesive ellipses depending on the language (English vs. German), the register, mode (written vs. spoken mode) and production type (original texts vs. translations).

From an intuitive perspective, the definitions and descriptions of the overarching categories of ellipses as will be specified in the annotation scheme do not give us a strong a priori reason to assume that either English or German are characterised globally by more instances of ellipses used as cohesive devices. In both languages, we find optional omission and reduction strategies which can occur in a great variety of discourse contexts with apparently no binding rules on when to use them or to opt for more explicit surface structures. Some general contrasts between English and German can be anticipated with regard to more fine-grained micro-categories that are described in the subsections of Chapter 5. In English and German, ellipses and fragments have similar functions. Ellipses are not extremely numerous cohesive devices and I expect them to occur a handful of times in most texts in both languages. In some texts, cohesive ellipses may appear abundantly while in others, they do not occur

at all. This will probably result in positively skewed frequency distributions for the individual texts. We may find some zero values, many values relatively close to zero and a long tail of the distribution towards the right.⁵

The body of existing literature reveals no strong hypotheses for the comparison of English and German with regard to their distribution of cohesive ellipses, but we can derive several indirect hypotheses from previous studies on related aspects. Specific linguistic differences on several levels between English and German have been assumed by Hawkins (1986: 28ff.), Leisi and Mair (2008: 65ff.), Leech et al. (2009: 20, 239), Mair (2006: 183), König and Gast (2012: 246ff.), Hansen-Schirra et al. (2012: 76ff.) and Fischer (2013). Most of the earlier assumptions are related to systemic contrasts between the English-German language pair and to differences with regard to specific lexical or grammatical properties. Only recently, frequency effects and cohesive patterns have received closer attention in contrastive studies.⁶ Many publications on cross-linguistic variation such as Hawkins' now-classic book on English-German contrasts (1986) cover typological aspects from a broad perspective and do not directly address discourse phenomena and textual relations such as ellipses-antecedent relations. From Hawkins' comparative typology of English and German, we can derive some general

⁵ It will probably be difficult, if not impossible, to transform the variables to arrive at approximately normal, bell-shaped distributions if specific variables have a substantial number of zero values.

⁶ So far, studies on discourse relations have mainly been concerned with certain cohesive devices in particular registers, but not with cross-linguistic difference or usage patterns across a wide range of written and spoken registers.

assumptions on ellipses. Hawkins assumes that English has drifted from a 'tight fit' language with originally richer, more complex surface forms that map onto less ambiguous meanings to a 'loose fit' language, while Modern German still is a relatively 'tight fit' language with more of a one-to-one correspondence between form and meaning (Hawkins, 1986: 121-127; 2014: 143-146). Hawkins' early work explained cross-linguistic differences with differences in inflectional morphology, but his later work rather emphasized the VO/OV parameter from the perspective of processing efficiency.

Based on this general difference, Hawkins expects more deletions of arguments from surface structures in English compared to German. This is believed to result in more deletions of entire noun phrases in English. Although Hawkins made no explicit statement on any other types of deletions, it may be possible that this linguistic difference also results in more deletions *within* phrases in English. The linguistic phenomena investigated by Hawkins led him to the general conclusion that German grammar is more explicit than English grammar.⁷ Hawkins mainly related his statements on deletions to sentence-internal phenomena, but the rules he formulated are likely to apply across sentences as well. My general hypothesis is therefore that English uses more omissions. I expect more ellipses of the cohesive and the non-cohesive type in English than in German.

⁷ Kortmann and Meyer (1992) raised doubts whether this is generally true. It has been acknowledged that English has certain grammatical compensatory strategies – that German lacks – to make up for the lower degree of explicitness in areas such as the English inflectional system, for instance, by showing a closer match between meaning and form in adverbs by explicitly marking this word class with '-ly'.

Another aspect, closely related to the issue of explicitness, is the level of grammatical complexity, which is equally difficult to measure exactly and to compare across languages (cf. Newmeyer and Laurel, 2014). One criterion for measuring complexity that has been suggested is the amount of irregularities in a language (McWhorter, 2008: 167).⁸ If we consider ellipsis contexts and fragments as irregular constructions that deviate from standard syntactic constructions, we can assume a (vague) relationship between the use of ellipses and fragments and the level of grammatical complexity of a language. McWhorter (ibid.) claimed that it is inherent to natural grammars to maintain a considerable level of complexity over time so that when simplification occurs in one area, this is usually counterbalanced by complexification in another area. Languages that are acquired natively by each generation usually do not drift into any overall simplification, but McWhorter sees English as an exception to this rule showing some grammar drift into relative, albeit not extreme, simplicity compared to other Indo-European languages.⁹ If we followed McWhorter's argumentation, we would expect German to have more ellipses (and other fragments, cf. Chapter 6) as irregular syntactic structures than English, but admittedly such a hypothesis based on the assumption of different levels of language complexity is not a strongly-

⁸ McWhorter's studies have a particular focus on the comparison of languages which have been restructured to a different extent by contact with other languages.

⁹ English is the only language that lacks various grammatical features which exist in other Germanic and Indo-European languages, and while English has developed some new complexities of its own (as other Germanic languages have also developed their own complexities), these do not approach the volume of the 'losses' in English (McWhorter, 2008: 168). McWhorter argues that the entirety of these losses is due to extensive non-native acquisition of English by Scandinavians during a period of intense language contact in a previous stage of English.

supported, uncontroversial claim. As said above, based on Hawkins' argumentation, I generally expect more deletions in English. From McWhorter's observations, on the other hand, we may derive the cautious assumption that German has more fragments as irregular syntactic structures.

Neumann's study (2013) presents an analysis that compares English and German on the basis of the CroCo corpus, a data set that contains the same texts that are also included in the written corpus sections of the current GECCo corpus. Neumann observed that English in general prefers recurrent units and that this may even apply to larger units and more abstract categories (2013: 316ff.). House (2011: 171) likewise had claimed that the preference of recurrent units and routine formulae in English texts is closely related to the distribution of cohesive devices in texts. Bisiada's corpus study (2013) on English and German equally showed that users of German prefer more situation-anchored, ad-hoc formulations than users of English, for example in marking causal relations. If it is true that English uses more recurrent units than German, this may have an effect on the use of ellipses, albeit not in a consistent direction. A high number of recurrent grammatical patterns and the frequent use of parallelisms in English can lead to a higher number of ellipses of certain elements in parallel structures. On the other hand, frequent repetitions of lexical material in English can correspond to a variety of strategies to avoid lexical reiterations in German, such as the use of ellipses, synonyms, general words, hyponyms etc. We can expect to find a certain relation between the

frequency of recurrent units in a language and the frequency of ellipses, but some types of recurrent patterns may lead to fewer ellipses, while others may have the opposite effect.

Some general contrasts between English and German can be anticipated with regard to the frequencies of nominal and verbal/clausal ellipsis as specific subtypes of ellipses (cf. Chapter 5). It has sometimes been claimed that English prefers a style with more verb forms than German and that English is more verb-oriented and encodes more information in verbs, while a more nominal style is more characteristic of German (cf. Kunz, 2010: 172). If English tends to use more and longer verb phrases, this probably also leads to more potential options than occur in German to leave out verb phrase elements or longer sequences in clauses. If German, on the other hand, prefers a nominal style, there should be more and longer noun phrases in German compared to English and at the same time more potential options for omission of head nouns. I initially assumed that we could expect to find more nominal ellipses in the German data compared to the English data due to the more nominal style of German and the fact that nominal ellipsis in English is sometimes blocked if a nominal substitute has to be used. On the other hand, I thought that we should expect that English will have more verbal ellipses due to the more verbal style of English and more options to leave out verbal element in texts that are rich in various types of verbal structures. Moreover, German sometimes has to use additional elements such as pronouns to circumvent the use of certain types of verbal ellipses that can be ungrammatical in

German. Nevertheless, a study on the CroCo corpus, GECCo's predecessor, showed that it is not a straightforward matter to determine which language is more noun- or verb-oriented. The level of nominal and verbal style for the CroCo corpus was investigated in Hansen-Schirra (2011). A closer look at results from this study made me revise my initial claims because, surprisingly, Hansen-Schirra found that the English and German corpus texts have similarly high values indicating a nominal style. Compared to German, the English data even have slightly higher values indicating a nominal style and slightly lower values than German indicating a verbal style (ibid: 145).¹⁰ This will probably also have an impact on the frequencies of the ellipsis subtypes in our data that are largely based on the textual data that were analysed by Hansen-Schirra for verbal vs. nominal style.

We expect differences between original texts and translated texts of the same language due to source-language interference. Toury (1995: 275) considers source-language interference as a translation universal or a 'translation law'.¹¹ English translations will probably differ from English originals and show some similarities to the corresponding German originals and vice versa. Although in certain cases, it will be difficult or impossible for translators to use an ellipsis or the same type of ellipsis as in an original text, translators will often try to follow the syntax of the

¹⁰ Hansen-Schirra (2011: 144) assumed that nominal style is a typical feature of content-oriented registers in both English and German while the use of pronouns and verbal style is characteristic of addressee-oriented registers in both languages. Most of our corpus registers seem to be strongly content-oriented.

¹¹ Others have claimed that source language interference contradicts universality in the translation process as it tied to the individual characteristics of languages.

English or German original text closely and to omit the same material in the translation that would otherwise have to be repeated in close proximity to a textual antecedent. In addition to such effects of source language interference, I would like to test the often-cited explicitation hypothesis as one of the proposed translation universals. Shifts in translation where certain cohesive devices from the original text tend to be replaced by different, more explicit cohesive devices in the target text were first addressed by Blum-Kulka (1986). If translations tend to become more explicit in the translation process, one would assume that translators systematically and often unconsciously replace pro-forms with content words and ellipses with complete structures. The notion of translation universals and the explicitation hypothesis have been a controversial subject for some time (e.g. Baker, 1993; Chesterman, 2004), but have not been fully confirmed or rejected by previous studies. Both explicating and implicating shifts were observed, for instance, in an English-German parallel corpus by Becher (2011).

I think that there may be a slight tendency to make abridged structures more explicit by adding lexical material in the translation process and by creating more explicit relations between textual elements. Therefore, I expect translations to have fewer ellipses than the original texts of the same language, but as the syntactic patterns in translations also reflect an influence of the source language, I expect translated texts to behave slightly similarly to their source texts.

Apart from the cross-linguistic differences and the differences between translations and original texts that I expect to affect the frequencies of ellipses in our English and German data, I assume that the use of the ellipsis types strongly depends on the register. It has already been found in contrastive studies on English and German that several contrasts in the use of co-reference, substitution and conjunction are more pronounced between different registers independent of the language (Kunz and Lapshinova-Koltunski, 2015). It is probable that we will see similar results with regard to ellipses. We can expect to find certain similarities in texts belonging to similar registers regardless of the language involved. Additionally, I would like to test the intuitive assumption that all types of cohesive ellipses tend to be indicators of spoken, dialogic and addressee-oriented registers and occur more frequently in such registers in both English and German. As most of our corpus registers consist of written texts of which many are strongly content-oriented, we will probably not find numerous cohesive ellipses in such texts, but many more in the spoken corpus sections and in those written texts that are very addressee-oriented or that include fictional dialogues.

The majority of *non-cohesive* ellipses in contrast to cohesive ellipses in our English and German data will probably be endophoric clause-internal ellipses or ellipses in coordinated structures and rather serve as indicators of written language.¹²

¹² Exophoric ellipses among non-cohesive ellipses would be indicators of spoken language, but they will be rare in our corpus data. In previous studies on co-reference we

Finally, I assume that German will show a globally stronger registerial differentiation and stronger distinctions along the dimensions of written vs. spoken and formal vs. colloquial style than English. The reasons for this assumption are complex and their finer details are beyond the scope of this dissertation.¹³ In short, I assume that the increasing influence of colloquial forms and a drift towards more oral linguistic characteristics in written language is an ongoing process which has become more persistent in English than in German. This phenomenon has been referred to as ‘colloquialisation’, ‘conversationalisation’, ‘vernacularisation’ or ‘informalisation’ and has been attributed to disappearing boundaries between public and private spheres in Western societies in general (Habermas, 1962: 262). It has also been attributed to the contemporary tendency in public communication and the mass media to model public discourse upon the discursive practices of ordinary life and to adopt conventions of one-to-one, face-to-face interaction between equals (Fairclough, 1994). In the last two decades, the effect of colloquialisation has been studied extensively on the basis of corpus data of English and its varieties around the world from both the ‘Inner Circle’ and the ‘Outer Circle’ (e.g. in Collins, 2012) and colloquialisation is believed to be characteristic of English.¹⁴ Originally, the assumption of colloquialisation

have already found that our corpus data do not include many exophoric co-reference items either. Therefore we can neglect the effect of exophoric ellipses for our data set.

¹³ They have been briefly addressed in Kunz et al. (forthcoming a) and Kunz et al. (forthcoming b).

¹⁴ Only a few researchers who observed similar tendencies of a colloquialisation of public discourse in other languages described it as a ‘global macro-phenomenon’ of language change (Doleschal and Hoffmann, 2004: 70). More details on the debate on colloquialisation can be found in Kortmann (2001) who provides an overview of the

was primarily related to the textual instances of lexical variation and the frequencies of grammatical indicators of informality and orality. Several studies have shown that colloquialisation also influences the use of cohesive patterns. Neumann's study of contrastive register variation (2013) presented a quantitative approach to the comparison of English and German written registers in the CroCo corpus and came to the conclusion that there are indeed more similarities among the English registers when compared to the German ones that are characterised by more cross-registerial diversity.¹⁵ I will test the hypothesis of a lower registerial

discussion on the drift towards more oral linguistic characteristics in written English and the colloquialisation of English written registers summarising Biber and Finegan's (1989, 1992) and Mair's (1997) studies. Biber and Finegan had analysed various written and speech-based English genres from the 17th to the 20th centuries. Mair had mainly investigated newspaper language from a diachronic perspective. These studies and Kortmann's observations confirmed a growing informality in English-speaking modern societies leading to the observation that written English, probably under the dominating influence of American English, increasingly allows for constructions that were more typical for spoken informal language in the past (e.g. increasing frequencies of quasi-modals, the progressive form, the *going-to* future and contractions like 'he'll', 'I've' or 'I'd' in written English, cf. also Mair, 1997: 203; Hundt and Mair, 1999; Mair, 2006: 183; Leisi and Mair, 2008: 65ff.; Leech et al., 2009: 20, 239 and Collins, 2012). We assume that in English this can result in more lexical verb ellipses in contexts of complex and long verb phrases containing quasi-modals, progressives or the *going-to* future. At the same time it will lead to fewer lexical verb ellipses in other contexts where auxiliaries or modal verbs undergo contraction. Mair acknowledged that, additionally to changes of the structural inventory of English that often originated in spoken language, the macrostructure of written English texts has changed over time due to the changes in discursive practices that Fairclough described, which Mair sees as an important driving force for further language change. Newspaper articles, for instance, now contain more direct quotations than in the past, and effective headlines and opening and closing paragraphs often display strategies of feigned orality (cf. the report on Mair's corpus linguistic project in Raible, 1998: 244). Mair came to the conclusion that it seems plausible to generalise his observations to written English as a whole (Mair, 1997: 203). Biber and Finegan (1997) stated that written English registers have undergone a split, with certain specialist expository registers steadily evolving towards more literate characteristics, while the majority of written registers indeed reverse course and shift back towards more oral styles. A stricter separation of highly specialised registers from general language is sometimes claimed of English compared to German (Leisi and Mair, 2008: 46ff.).

¹⁵ For translations this depends on the text type. It has been found, for example, that routine formulae that are typical of certain English genres such as children's literature are often omitted or replaced by non-formulaic, ad-hoc forms in German translations of English texts (House, 2004). This also depends on the language proficiency of the individual translator as there seems to be a tendency in non-native speakers and learners

differentiation in English with regard to the use of ellipses and a less sharp distinction between written and spoken mode in English compared to the German data.

To sum up, I expect English to have more cohesive ellipses in total, but I may observe different results with regard to the frequencies of certain ellipsis subtypes. German probably has higher frequencies of other types of fragments as irregular syntactic structures. Translations are expected to behave differently compared to original texts from the same language and also to show some similarities with their source texts. Possibly, a considerable proportion of ellipses are translated by the use of more explicit structures. I expect the frequencies of ellipses to depend mainly on the register being instantiated by a text. The contrasts between groups of texts belonging to different registers may be more pronounced than differences between English and German groups of texts belonging to the same register. I assume that German will show a globally stronger registerial differentiation. Substantial differences can be expected between the written and spoken modes, although specific texts within our written and spoken data may be characterised by extreme values compared to the mean values of our written and spoken subcorpora. All types of cohesive ellipses probably tend to be indicators of spoken, dialogic and addressee-oriented registers and occur more frequently in such registers in both English and German. I expect to find certain cross-linguistic similarities between the registers regardless of the language involved.

towards more ad-hoc formulations as Barron (2003: 181) suggested in her study on Irish learners of German.

2. A brief historiography of ellipsis as a grammatical concept

An extensive body of literature on ellipsis has accumulated over time, and it is important to trace the emergence and historical development of ‘ellipsis’ as a linguistic concept to explain some of the reasons behind the complexity and loose definitions of this term. The diverse literature on ellipsis has resulted in some confusion over what exactly ellipsis is. That is why modern linguists have labelled ellipsis a ‘problem’ (Ortner, 1987), a ‘category mistake’ (Buss, 2004), a ‘stepchild in linguistics’ (Hennig, 2010: 76) or even a ‘plague’ (Bühler, 1990 [1934]: 189).

To a certain extent, the linguistic concepts and categories linguists now use to analyse and describe languages such as English and German have been shaped by the heritage of very early grammatical, philological and rhetorical discourses. The historiography of linguistic concepts is not unimportant or negligible, as it helps to explain how certain definitions have evolved that are now applied in contemporary linguistic studies and put forward in pedagogical grammars. Ellipsis is a rather complex linguistic concept. It is related to concepts such as ‘word’, ‘sentence’ or ‘cohesion’ which would be equally difficult to sum up in one-sentence dictionary definitions. Putting the concept of ellipsis into a wider historical context of Greek and Latin grammatical and rhetorical terminology and its reception and adaptation from the Renaissance to modern times is a useful step that will help to untangle some of the complex issues regarding ellipses. It will also contribute to an explanation for some of the reasons

behind the heterogeneity of attitudes and acceptability judgments as well as the relatively loose or even contradictory definitions of ellipsis that can be found in the literature.

A term used both in rhetoric and syntax, ellipsis has always had a dual nature as linguistic deviation and rhetorical figure at the same time. English and German as well as other modern languages borrowed the grammatical sense of the word ‘ellipsis’ from Greek – presumably this happened slightly later than the earlier recorded borrowing of the geometrical sense of the word, but it may have been at the beginning of the 17th century, at a time when more and more educational institutions were being established in Europe. The legacy of Greek and Latin grammatical categorisation and the grammatical thinking of seventeenth- and eighteenth-century prescriptivists from a particularly productive period in the development of grammatical norms and classifications are two very important factors that paved the way for our modern grammatical terminology and understanding.

The term ‘ellipsis’ comes from the Greek word ‘ἐλλείπω’ which means: *to leave out, fall short, fail, be in want of, fall short of, lack, be inferior to*¹⁶. Grammarians started to adopt this term from classical rhetorical and philological theory by analogy with ellipsis seen as a rhetorical strategy used for linguistic brevity, conciseness and focus. The Greek term for ellipsis then came into Latin usage about 2000 years ago. From the very beginning of the discussion on ellipsis, there has been some

¹⁶ Online Liddell-Scott-Jones Greek-English Lexicon:
<http://stephanus.tlg.uci.edu/ljsj/#eid=1&context=ljsj> [14/07/2016]

disagreement among scholars about its exact definition, its possible subtypes and the clear distinction from other structures. Some scholars would, for instance, have considered phenomena such as asyndeton, i.e. the absence of an explicit conjunction (e.g. *veni, vidi, vici* - *I came, I saw, I conquered*) or aposiopesis (an interruption or breaking off in mid-sentence) either as a kind of ellipsis or a separate and distinct phenomenon.

Nowadays, the broad and general meaning of the grammatical term ‘ellipsis’ according to the OED is the ‘omission of one or more words in a sentence which would be needed to complete the grammatical construction or fully to express the sense; *concr.* an instance of such a semantic or grammatical omission’. The Duden, a German normative dictionary, gives a similarly broad definition for ‘Ellipse’ as “Ersparung von Redeteilen” (omission of parts of an utterance) or “Satz, in dem Redeteile erspart werden; Auslassungssatz” (sentence involving an omission) where the nature of the unsaid elements is not specified.¹⁷

I observed that linguistic papers or pedagogical textbooks sometimes restrict their coverage of ellipsis to the everyday definitions found in such dictionaries, a method which can create a superficial picture of linguistic

¹⁷ The Duden gives a formulaic expression as an example of ellipsis that actually may have been perceived as an omission in the past, but has become a conventionalised phraseological unit: [*ich*] *danke schön* ([*I*] *thank you very much*). In another entry in the same source, the word *danke* in exactly this phrase is classified as a particle which shows that it must have undergone a grammaticalisation process and is not a typical case of ellipsis in modern German. cf. <http://www.duden.de/rechtschreibung/Ellipse> & <http://www.duden.de/suchen/dudenonline/danke> [last checked 15/07/2016]

phenomena.

Modern linguists continue to refer to ellipsis and its roots in antiquity although its meaning has constantly been extended and modified. Our modern understanding of English and German grammatical concepts such as ellipsis is, at least indirectly, influenced by grammatical descriptions of classical languages, but I would also like to emphasise the influence of the writings of early European grammarians such as seventeenth- and eighteenth-century prescriptivists, who extensively discussed ellipses and the notion of the complete and well-formed sentence in debates about correct language use. From the middle of the 18th century on, numerous prescriptivist and normative grammars were published in German and in English. Ellipsis in the context of the discussion of the complete sentence as the underlying unit became an increasingly discussed topic in English and German linguistics, but it remained an ill-defined term as, for instance, a list by Schiefer (1974: 210-213) with numerous definitions of ellipses by 18th and 19th century authors with regard to various languages shows.

The authoritarian and normative nature of 18th-century English grammars gave rise to a 'doctrine of correctness' (Leonard, 1929) according to which expressions and syntactic structures such as elliptical sentences were seen as either correct or incorrect. Finegan (1998: 547) claims that English prescriptive grammarians from this period were heavily influenced by Latin grammatical concepts and categories, and often attempted to make English grammar follow Latin rules. There has always been a relatively subjective acceptance and rejection of certain

grammatical structures, including elliptical constructions, particularly by grammar prescriptivists. In an attempt to standardise and stabilise the English language and to reduce it to a certain set of teachable rules, prescriptivist grammarians started to set up strict rules of grammatical usage condemning those forms and constructions that they considered to be improper. Smith (1986: 73) and Sundby et al. (1991: 242) ascertained that the concept of ellipsis varied during the 17th and 18th century and largely depended on the professional judgment by individual grammarians. Both scholars pointed out that elliptical constructions were extensively discussed by 18th century English grammarians who made a distinction between omissions as a grammatical figure and a grammatical error – as Greek and Latin grammarians had already done a long time ago.

It is possible that particularly English still reflects a strong influence of the grammatical heritage and analytical methods of early Latin and Greek grammarians that were enjoying a revival in a period of time when the standardisation of English received much attention. A standard ‘fixed’ variety including strongly standardised syntactical rules began to be developed in the 18th century for a range of functions while non-standard use became more associated with simplicity and informal purposes such as the language at the market or at home. Fewer deviations from standard phrasal structures and sentence patterns were tolerated from then on in English. This may explain a tendency to see ellipses and fragments as inferior compared to other structures, which is probably more pronounced in English than with regard to German. With regard to English, we may

expect fewer deviations from standard phrasal structures and sentence patterns.

Highly depending on the definition of a basic linguistic reference unit, ‘ellipsis’ is a relative concept as it refers to what is understood as a correct or complete proposition or sentence. However, the sentence is another complex concept whose meaning may seem evasive (cf. for example Morgan, 1973; Forsgren, 1992; or Dürscheid and Schneider’s 2015 handbook on the concepts of sentences and utterances). The current debates on the meaning of ‘sentence’ are dominated by approaches and definitions that occurred from the 1950s onwards. Syntactic research in the last sixty years deeply influences recent linguistic research as Graffi’s “200 Years of Syntax – A critical survey” (2001) shows. However, in the chapter “The analyses of the sentence and of the word groups”, a very detailed overview of the development of the sentence concept, Graffi (2001: 111-166) demonstrates that syntactic matters and the definition of ‘sentence’ were in fact carefully investigated long before, just as the related concept of ellipsis has been a topic for discussion for a long time. Starting with an examination of definitions of ‘sentence’ which classical tradition handed down to us (e.g. by Priscian who defined a sentence as an arrangement of words that follows some criteria of combination having the capacity of expressing a complete thought), Graffi continues to discuss the Port-Royal grammarians’ tripartite structure of the sentence (which they called a ‘judgment’): Subject, Copula and Predicate. All syntactic theories

had and still have to struggle with sentence types and utterances that do not seem to fit into the respective model.¹⁸

Gabelentz (1874-75: 141f.), Sweet (1875-76), and Wegener (1885) had already called attention to word sentences ('Wortsätze'), sentence words ('Satzwörter') and one-member sentences ('eingliedrige Sätze') that some have called the archetype of human speech (cf. Berka, 1999: 111, referring to Gottlob Frege's writings).¹⁹ This issue of subsentential speech is not a new one and in recent time it has been emphasised by several authors that genuinely subsentential phrases and non-sentential speech can be used to perform speech acts without being cases of syntactic ellipsis (e.g. Stainton, 2004 *inter alia*; Scheffler, 2005; Hall, 2009).²⁰

¹⁸ The Port-Royal model of sentence analysis already had difficulties in explaining "real" verbless sentences, i.e. sentences which contain only a subject and a non-verbal predicate, e.g. Latin *'Vox populi vox dei'* (People's voice <is> God's voice) that seem to be a minor sentence type in many Indo-European languages and must be kept distinct from ellipsis in the sense of gaps in incomplete and imperfect sentences (Graffi, 2001: 131f.).

¹⁹ Lugebil (1884: 60–67, cited in Graffi, 2001: 115), for instance, discussed 'epiphonematic expressions' (e.g. *'What a shame!'*), infinitives and participles with a command meaning (e.g. *'Come here!'*) or absolute infinitives like *'so to speak'*.

²⁰ In his study, Graffi (2001) also focusses on research by German linguists and philosophers throughout the 19th century and during the first half of the 20th century, listing sentence definitions by Wilhelm von Humboldt, Karl Ferdinand Becker, Karl Wilhelm Ludwig Heyse, Heymann Steinthal, Johann Friedrich Herbart and several others. He refers to Ries (1931) who listed more than 150 sentence definitions, most of them from the middle of the 19th century onwards and from the beginning of the 20th century. Some, for instance, distinguished between the linguistic sentence as a concrete instantiation of the abstract structure of 'judgment' or called the sentence "an expression of thought" or "an uttered thought" formed by a Subject and a Predicate. As a consequence of recent evolving and competing theories, traditional definitions were rejected and new ones were proposed continuously. Graffi addresses the successive stages of the crisis of the Port Royal version of the 'judgment model' of sentence analysis and logic-based syntax (Graffi, 2001: 112) what resulted in the "divorce of grammar from logic" (ibid.: 24) and the abandonment of general grammar by historical-comparative linguists of the 19th century. He examines the rise and fall of theories such as 19th century 'psychologistic syntax' (associated with names such as Wilhelm Wundt or Heymann Steinthal) and shows that newer definitions often tried to avoid references to earlier grammar traditions – although they were not entirely free from them and many of the categories of older models were difficult to replace adequately (ibid.: 15). For instance, Jespersen coined the label of 'nexus' for the subject-predicate grammatical structure and tried to avoid any particular commitment to psychologism when defining

Due to the conceptual confusion between anaphoric elliptical omissions and other reduction strategies and principles of brevity such as subsentential speech and the fact that ‘ellipses’, ‘fragments’ and ‘sentence fragments’ are frequently used interchangeably as synonymous pairs, particularly in the English literature, many pedagogical grammar books for native speakers and foreign learners as well as stylistic guidebooks will advise the reader to avoid the usage of ellipses or fragments altogether, since they appear to reflect poor writing skills. Students can find numerous exercises to learn how to correct or ‘repair’ ellipses and fragments and to expand them into complete sentences. If pedagogical books recommend using ellipses and fragments in English writing at all, they usually emphasize that those constructions should be used intentionally, but sparingly: “Ninety-nine percent of the time you should use complete sentences in your writing” (Provost, 1988: 62). It is easy to find numerous similar pejorative judgments with regard to ellipses and fragments in German in prescriptive grammars and in books and articles containing writing and communication advice for native speakers and learners of German (e.g. Bubenheimer, 2001; Hofmann, 2011). Nevertheless, German style guides sometimes intuitively seem less strict with their recommendations to absolutely avoid fragments in writing and rather highlight that they should mainly be avoided in formal or academic writing or in translations to avoid ambiguity (Lehrndorfer, 2013: 671). In the

the sentence on the basis of its communicative function: “*A sentence is a (relatively) complete and independent unit of communication*” (Jespersen, 1933: 106).

Duden, we find the general recommendation to be careful when using ellipses: ‘[...] bei Auslassungen (Ellipsen) [...] sollte man Vorsicht walten lassen.’²¹

The challenges of clearly distinguishing ellipses from different structures are, in a way, similar to any early or modern attempts of putting linguistic elements into syntactic, semantic and stylistic categories. It would be interesting to trace the main stages of the early history of grammar writing in Europe in more detail, but this chapter confined itself to some aspects of the grammatical heritage of classical antiquity and the past centuries demonstrating briefly the complexity of the evolution of grammatical norms and of ellipsis as a grammatical concept. Due to space constraints, it is impossible to mention all historically relevant facts and to describe the historiography of ellipsis in great detail in this section. The reader is referred to Menzel (2016a) for a thorough description. The next chapter will have a strong focus on more recent literature on ellipsis in English and German to serve as the main input for developing a conceptual framework for the comparative analysis of ellipses in an English-German corpus.

²¹ <http://www.duden.de/sprachwissen/sprachratgeber/nicht-zu-empfehlende-ellipsen> [last checked 15/07/2016]

3. The description and analysis of elliptical phenomena

3.1 Ellipsis and the written-spoken dimension

The previous chapter contended that it is worth looking at how the term ‘ellipsis’ has evolved from classical antiquity to the early modern period while this chapter has its focus on the conceptualisation of ellipsis, changing linguistic theories and methodological prerequisites in Anglophone and Germanophone academic discourse from the more recent past. It discusses methodological and analytical, qualitative approaches that have been covered in the last decades with regard to ellipsis and will address several central and general questions related to the research on ellipses as well as some long-standing controversies.

The categorization of ellipses is still a prevailing issue due to the ‘written language bias’ (Linell, 2005) which continues to affect our conceptualization of language models and the understanding of the sentence concept. This traditional dominant social perception of written language as a standard or ideal of language is one reason for the sometimes pejorative treatment of ellipses. This discussion is related to the issue of determining where one utterance ends and the next begins. A thorough summary and comparison of several approaches of discourse segmentation with regard to written and spoken language can be found in Stein (2003) who shows that this is not a theory-neutral activity.

Possible ‘complete’ structures of patterns which are usually analysed as ellipses are particularly connected to the normative character of written language use, and many conventional descriptive grammar books focus primarily on the description of standardised language that can be used in written texts while usually not recognising a need to address particular phenomena from spoken discourse. Nevertheless, the discussion on ellipsis has become more complex with growing interest in the description of the specific syntax and discourse structures of spoken language. Interest in spoken language has gradually intensified from the 1960s onwards. Additionally, the introduction of speech-act theory and discourse analysis has led to a greater interest in the communicative function of the typical linguistic devices of spoken language. Particularly within the last few decades, several authors have started to develop comprehensive theories to describe and analyse the nature of spoken English or German from a grammatical and discourse perspective, such as Brazil (1995), Schlobinski (1997, ed.)²², Jürgens (1999), Biber et al. (1999), Sandig (2000), Schwitalla (2003), Deppermann (2006), Hennig (2006), Ágel and Hennig (2007) and Auer (2007).

Earlier studies on spoken language were mainly contrastive in nature and oriented towards the syntax of spoken language only in the sense of comparing it to the syntax of written language. In those studies, spoken language was often considered a derivative linguistic system that deviated from written language and had its own rules. It was sometimes even seen

²² cf. the chapter on ellipses in spoken language by Selting (1997).

as a deficient, erroneous linguistic form of language. The traditional dichotomy between spoken and written language considers written language generally as rather “artificial” (cf. Ong, 1982: 82), monologic, planned, elegant, formal, clear, complete and correct and spoken language as natural, dialogic, spontaneous, sloppy, familiar, incomplete and incorrect.

Among the characteristic features of spoken language that have been pointed out (for instance in Bussmann, 1996: 1115) are anacoluthons (syntactic blends), specific syntactic structures such as left vs. right dislocation, hanging topics etc., dominance of paratactic over hypotactic structures, a frequent use of discourse particles and discourse markers and a high frequency of short and often incomplete sentences (‘ellipses’ or rather ‘fragments’ as I will call such cases of non-clausal units, cf. Chapter 6). While it has often been claimed that incomplete sentences are characteristic of spoken language, endophoric ellipsis and cross-sentential ellipsis-antecedent relations have been less frequently addressed and it is not clear whether they are a typical feature of written or spoken language.

It is necessary to specify what is meant by the underlying segments or building blocks of various discourse contexts and if one wants to make a distinction between the grammar of written and spoken language. Identifying sentences or discourse units in spoken language is particularly difficult and it has been discussed whether the sentence as the primary unit of language is adequate for capturing the whole picture of spoken language (cf. Stein, 2003). Some linguists deny this (e.g. Rath, 1976; 1985: 1653;

Fiehler et al., 2004: 181) and claim that interlocutors use specific signals for structuring spoken conversation. Lexical and syntactic signals co-occur with prosodic signals and pauses. It is possible that these kinds of structuring devices are generally incompatible with sentence units. Betten postulates that there is no strict rule that makes speaking in full sentences necessary: As soon as you realise that the hearer has understood the information you wanted to convey, it is unnecessary or even inappropriate to continue talking (Betten, 1976: 225). Betten came to the conclusion that it is this principle that results in fragmentary speech. This principle is related to Grice's Maxims of Conversation (Grice, 1975). The Maxims of Manner and Quantity imply that it is advantageous to omit elements which are retrievable from the context.

Koch and Oesterreicher (1985: 25) see different types of discourse on a scale from communicative immediacy to communicative distance and between 'conceptually oral' language (interactive texts, language as an active process – often, but not necessarily, in oral mode) and 'conceptually written' language (i.e. highly planned and edited texts, language as something created and produced – often, but not necessarily, in written mode, cf. also Raible [2001: 2]). Koch and Oesterreicher's model of conceptually spoken and written style separates the phonic or graphic realization from 'conception' that can be spoken or written style (cf. Söll, 1974). It relates linguistic features to communicative situations by means of a continuum between nearness and distance, based on the level of spontaneity, formality, monologue vs. dialogue, synchronous vs.

asynchronous mode and a continuum between written and spoken style, based on syntactic and lexical features and discourse structure.

There are typical conceptually written and spoken registers using compressed language requiring familiarity with their conventions. In certain registers, information has to be presented in a concise way due to principles of language economy or to be memorable. We can therefore expect ellipsis to occur in both conceptually written and spoken registers depending on more complex variables than the written-spoken dimension. In this present study, no particular difference will be made with regard to utterance segmentation in written and spoken language. Punctuation to mark sentence boundaries is assumed to be correct in the written part of the corpus and the transcriptions of the spoken part which were generated by human transcribers according to the guidelines of the project. Treating the written and spoken corpus registers within the same grammatical framework also makes sense because the distinction between written and spoken language is relatively gradual in the registers of the GECCo corpus which we could place on a written-spoken continuum in the sense of Koch and Oesterreicher's model despite some general differences between written and spoken mode. There are written registers in GECCo with relatively oral situational characteristics (cf. also Biber, 1988: 45), such as letters to shareholders or prepared speeches, and spoken registers with relatively literate situational characteristics such as academic discourse or sermons.

3.2 Ellipsis and the syntax-semantics-pragmatics interface

One topic that continues to be a focus of interest with regard to ellipsis and that is related to the written-spoken distinction but also to the understanding of the sentence concept discussed in Chapter 2.3 is the syntax-semantics-pragmatics interface. Different camps can be distinguished that give different answers to the question of whether ellipsis resolution is mainly syntactic, semantic or even pragmatic in nature and whether ellipsis involves deletion or not.

A well-established branch in modern linguistics that has been extremely productive on ellipsis is transformational generative grammar, starting with Chomsky (1955, 1965). Many publications followed and examined various types of anaphoric relations (e.g. Wasow, 1972; 1979; Sag and Hankamer, 1984, Chao 1987) and ellipsis, particularly in the generative grammar tradition (cf. also van Craenenbroeck van Merchant's description of ellipsis phenomena in a handbook of generative syntax [2013]). This branch contributed a great deal to the debate on the notion of identity between the antecedent and target of deletion. Scholars such as Lyn Frazier, Daniel Hardt, and Jason Merchant, to name but a few, provided numerous publications on the structure of ellipses. Scholars such as Arregui et al., 2006; Duffield and Ayumi, 2009; Frazier and Clifton, 1998, 2005, 2006 inter alia; Garnham and Oakhill, 1987; Kertz, 2010a/b; Kim and Runner, 2011; Lappin (1992), Lappin and Shih (1996), Martin and McElree, 2009, inter alia, investigated ellipsis processing and resolution.

Winkler (2005) provides a detailed summary of research on the ellipsis concept in the generative grammar tradition.

As ellipses have often been thought of as incomplete sentences, we can choose to look at them from an either more syntactic, semantic or pragmatic angle. More generally, as mentioned before, an ellipsis can be understood as an omission of an element in an underlying, abstract and idealised complete reference unit, but it is usually claimed that this reference unit is the grammatically correct and complete sentence, clause or phrase.

Nowadays, linguists typically refer to at least three senses of the term ‘sentence’ (Stainton, 2006a: 31):

- a. sentence_{syntactic}: an expression with a certain kind of structure/form,
- b. sentence_{semantic}: an expression with a certain kind of content/meaning,
- c. sentence_{pragmatic}: an expression with a certain kind of use.

The sentence concept today is defined in syntactic, semantic and pragmatic terms, but also on the graphic and prosodic level. It include factors such as speech prosody or the falling intonation which we naturally associate with the end of an utterance in spoken language and the punctuation used to mark the closure of an orthographic sentence in written language.

Ellipsis has raised a number of questions on how syntax and semantics, but also pragmatics and phonology, interact with each other. Smith (2001: 176) states:

[Ellipsis] “remains difficult to classify, as it appears to involve phonology (due to its similarity to deaccenting), syntax (by virtue of its distribution), semantics (evidenced by its apparent licensing conditions), and pragmatics (because of the cognitive load it imposes).”

Phillips and Parker (2014) briefly summarise several studies that have focused on the relationship between ellipsis and antecedent and on the question of whether the antecedent of an anaphoric ellipsis is syntactic or semantic in nature. A central question concerns the content of the ellipsis site and whether or not there is unpronounced material, i.e. whether the ellipsis site hosts a detailed structural representation of the antecedent. The syntactic approach emphasises that there is some form of structural identity between the antecedent and the ellipsis site while the semantic approach emphasises the identity of meaning between them.

Psycholinguistics research has often addressed questions about the representation of sentences involving ellipses and suggested two explanations for analysing endophoric ellipsis and other types of ellipsis: the (syntactic) Deletion Hypothesis (e.g. Sag, 1976) and the (semantic) Interpretation Theory (e.g. Williams, 1977).

Various psycholinguistic studies and some studies of language acquisition have discussed whether the size of the ellipsis antecedent, the distance of the ellipsis to its antecedent or possible ambiguity have an impact on the timing of ellipsis resolution (Frazier and Clifton, 2000,

2001; Martin and McElree, 2008, 2009, 2011). If the antecedent – along with its semantic content and syntactic structure – is copied into the ellipsis site, as Frazier and Clifton (2000) suggested, it may be assumed that the copying operation induces higher processing costs for more complex antecedents, as more structure needs to be copied (Poirier et al., 2010). However, several studies have reported that the complexity of the antecedent and copying more structure do not necessarily cost more processing effort than copying less structure. In fact, the same number of inferences may be required independent of the amount of structure built (Martin and McElree, 2008; Frazier and Clifton, 2001: 2).

Emmott (1997: 221-235) claimed that, as a cognitive process, anaphoric elliptical reference is *forward-oriented*, i.e. language users make mental representations of recently mentioned information in anticipation of more implicit ‘references’, such as ellipsis. By contrast, standard theories of elliptical gaps as a type of *backward-acting* anaphoric reference are considered to be more appropriate (Wilson, 2000:12). Martin and McElree (2008) claimed that anaphoric verb-phrase ellipsis does not involve syntactic reconstruction at all; instead, representations formed during comprehension are content-addressable. Therefore, a content-addressable pointer mechanism directs the processor to the mnemonic representation of the antecedent. Additionally, Martin and McElree (2011) suggested that the distance between antecedent and ellipsis site affects the likelihood of successful comprehension, but not the time it takes to retrieve and interpret the antecedent (cf. Martin and McElree, 2011; Poirier et al., 2010). Xiang

et al. (2014) observed that exposure to endophoric verb phrase ellipsis in English affects syntactic choices in subsequent sentence production so that speakers favour the production of parallel structures. They concluded that ellipsis sites induce structural priming effects, which means that ellipsis sites contain or trigger access to a syntactic structure.

Culicover and Jackendoff (2005: 21) have pointed out that theories that are too ‘syntactocentric’ may be a barrier to fully understanding phenomena such as ellipsis and claim that syntactocentric thinking will necessarily lead to an interpretation of sentence fragments in terms of deletion or empty structures. Syntactic and semantic approaches have often competed with each other to explain ellipsis mechanisms; however, not only the syntactic, but also the semantic approach raise certain unanswered questions. Several authors have therefore emphasised the complex interplay of syntactic and semantic constraints with regard to ellipsis resolution.

One of the assumed benefits of elliptical spoken language in both the syntactic and the semantic view is a general reduction of the working memory capacity for planning the detailed shape of one or several constituents in an utterance, a hypothesis tested, for example, by Kolk and Hofstede (1994) and Kolk (2001). They provide a comparison between ‘normal’ elliptical speech, ellipses produced by young children and ellipses due to chronic agrammatic aphasia²³ examining whether an

²³ There are different types of aphasia. In some of them, comprehension is preserved and patients sometimes know what they want to say but cannot express themselves, which results in agrammatic sentence production. In other cases of sensory aphasia, speech is

overuse of the normal elliptical repertoire has the purpose of preventing the ‘computational overload’ that would result if a complete sentence form were attempted (cf. also Kolk and Heeschen, 1992: 89) and to ‘ease the communicative load’ (Penn and Cleary, 1988: 10) – in a way similar to shortening strategies in a telegram where each word has a cost (cf. also Tesak and Code, 2008: 131; Tesak and Dittmann, 1991).

Recent years have seen an increased focus on ellipsis from a pragmatic and discourse-analytical perspective. In pragmatic terms, resolution of the intended meaning of fragments can be modelled as a by-product of the establishment of discourse coherence and the establishment of certain connections of the content of the current utterance to the content of its discourse context. Pragmatic and prosodic aspects of ellipses in theoretical linguistics and the interplay of these factors with other levels have been addressed, for instance, by Carlson (2001) and Merchant (2010a).

In the syntactic approach, ellipsis is considered to be synonymous with the omission of constituents, i.e., the derivation from a complete sentence where fragmentary constructions are understood via complete syntactic deep structures (Fiengo and May, 1994; Merchant, 2001, *inter alia*).

So-called ‘omissionists’ (a notion used by Ortner, 1987 [‘Omissionisten’ / ‘Ableiter’]) consider fragmentary constructions as

preserved but comprehension is poor, and language content incorrect, so that non-existent or incorrect words may be used in fluent sentence production, e.g. “telephone” for “television”, and paragrammatic errors may occur. Agrammatism typically involves reductions in grammatical structures, but also confused, erroneous syntax. Neurolinguists such as Schlenck et al. (1995) or Ruiter (2008), Ruiter et al. (2010) claim that stimulating and automatising the production of ‘normal’ ellipses and telegraphic style in chronically agrammatic speakers increases their communicative efficacy and efficiency (Reduced Syntax Therapy).

deficient surface structures that are understood via complete syntactic deep structures, which is the traditional view. In this approach, ellipsis is regarded as a “phenomenon of overt syntax involving deletion” (Smith, 2001: 1). The deletion hypothesis goes back to the concept of the ideal grammatical sentence. Semantically, there can hardly be a definition of completeness of a proposition, because every utterance will only convey certain and not all aspects of a situation. There is still some discussion on the definition of syntactic completeness, but when compared with semantic completeness, it may seem easier to define elements that are structurally and grammatically necessary. For instance, a sentence, or more specifically a clause, can be considered as the smallest independent grammatical unit. Particularly in pedagogical grammar books, it has often been claimed that a grammatically complete clause includes a subject and a finite verb – two constituents at least and a sentence can consist of a single independent clause or a clause-complex.

Proponents of a more semantic or pragmatic account see ellipsis either as the result of semantic redundancy (e.g. Dalrymple et al., 1991; Dalrymple, 2004; Hardt, 1992a; Shieber et al., 1996) or as self-contained, autonomous structures. In the semantic approach, the interpretation of ellipses does not necessarily rely on underlying syntactic structures. They are interpreted in a process of establishing that a discourse is semantically coherent (cf. Williams, 1977, Kehler, 1993). Linguists who tackle this issue more from a discourse-grammatical angle have also emphasised that certain structures traditionally subsumed under ‘ellipsis’ are autonomous

constructions, a category *sui generis* or ‘verfestigte Ergebnisse kommunikativer Tätigkeit’ (Ortner, 1987: 132). Apart from Ortner (1985, 1987), several other linguists such as Klein (1993), Buss (2004), Schlangen (2003) and Corr (2010) discuss and defend this ‘what you see is what you get’ view.

Authors such as Barton (1990) or Progovac (2006) suggested a non-sentential constituent analysis arguing that non-sentential elements do not necessarily derive from a sentential source truncated by the operation of deletion rules. Zebrowska (2005) lists Eugen Lerch, Eugen Seidel, Rudolf Grosse, Wolfgang Mentrup and Franz Simmler as ellipsis autonomists who have warned against describing non-verbal / non-clausal structures against the background of verbal clauses. Carter and McCarthy (2006: 181) claim that if we take spoken language as the foundation of linguistic analysis we may come to the conclusion that, in reality, nothing is missing from elliptical messages and it makes equal sense to say that writing and formal speech need to elaborate more and add items that are unnecessary in everyday speech.

If ellipsis is defined as an omission and leaving something out, then the perception of what counts as an omission may theoretically involve subjective judgment, and the perspective of what a person was supposed to write or say might also change over time. Someone might have deliberately ‘omitted’ a structure or simply not written or uttered it as a form of non-realization, i.e. the optional absence of potential elements from a sentence. But in many cases it remains unclear whether we can call

it an ellipsis if it did not even occur to a person to express a construction in another way. Generally speaking, autonomists have a stronger focus on a discourse-analytical approach describing syntactic regularities involved in the production of various fragment constructions and explaining their communicative function.

Autonomists, in contrast to omissionists, argue that we find patterns where the relation to the co-text and context remains relatively open – where the syntactic, semantic and discourse connection is unclear or ambiguous so that it is difficult to paraphrase and to complete them. This view seems to account for a wider range of non-sentential utterances, particularly with regard to spoken language. Autonomists point out that not every non-clausal fragment or unit smaller than a sentence should be analysed as being the result of an omission or actually needs to be interpreted as the reduced version of an underlying complete sentence.

Paul (1995 [1880]: 313) already stated:

‘Misst man allemal den knapperen Ausdruck an dem daneben möglichen umständlicheren, so kann man mit der Annahme von Ellipsen fast ins Unbegrenzte gehen. Bekannt ist der Missbrauch, der damit im 16. und 17. Jahrhundert getrieben ist.’

This means that one can theoretically read into every isolated word or expression an unexpressed, implicit linguistic environment, something that

was apparently done extensively during the 16th and 17th centuries. Bühler's recommendation for an "effective and radical cure for the ellipsis plague that has been with us now for two millennia" (Bühler, 1990 [1934]: 189²⁴) was to limit the flood of constructions that might fall under the term 'ellipses' and by not seeing an implicit or unexpressed linguistic environment in every isolated occurrence of a word or expression as this might be an arbitrary and overgeneralised application of the concept of ellipsis (cf. Van Ginneken's [1910] and Jespersen's warning against "ellipsomania" [1937: 167], Busler and Schlobinski [1997] and Feilke et al. [2001: 5] on "ellipsophile" grammarians).

Autonomists have questioned whether it always makes sense to take assumed complete structures from which certain elements have been omitted as the starting point of the analysis. Additionally, they questioned whether such structures that are called 'ellipses' or 'fragments' can always be completed in an obvious and unambiguous way by adding more syntactic material or whether this would only be an arbitrary way of paraphrasing those structures.

However, in the case of endophoric ellipsis, on which I will mainly focus in this thesis, it seems legitimate and correct to speak of deletions or omissions. What is important for the classification of our ellipsis categories is the underlying sentence or phrase with a complete syntactic or phrasal structure assuming that there is unpronounced material.

²⁴ The pagination of the German source text (1934) differs, but the corresponding number is indicated on the left side of the translated text in the English version.

Autonomous non-sentential constructions and other fragments will be subsumed under a separate category in Chapter 6.

3.3 Ellipsis scope, non-identity effects and acceptability

Apart from the general discussion about ellipsis in syntactic, semantic or pragmatic accounts, ellipses resulting in ambiguity and underspecified scope where the context has a significant influence on the interpretation of elliptical sentences have raised much interest and have sometimes been taken as evidence for semantic processing. The following cases of unclear scope have been discussed, for example, in Scheffler (2005: 5ff.) who pointed out their problematic nature for both syntactic and semantic accounts of ellipsis:

- Scope Parallelism:

(3:1) *Lilly introduced exactly one student to every girl, and Mimi*

did []²⁵, too. [] = 'introduced exactly one student to every girl'

The elliptical verb phrase contains two quantifiers and since there are two scopal possibilities for two quantifiers, one would theoretically expect four different possible readings for the whole sentence. Instead we observe that the reading of the second clause is determined by the scope reading of the first one. “Only two of the four possibilities are grammatical, because

²⁵ In the examples here and below, an ellipsis site is indicated by '[]'. '[/]' indicates that there is no underlying syntactic structure and '[?]' is used for cases where it is debatable whether there is an ellipsis site or not.

the semantics of the elided clause depends on the semantics of the overt clause” (Scheffler, 2005: 6).

- Hirschbühler's puzzle:

(3:2) *A Canadian flag was hanging in front of each window, and an American one was [] too.*

Hirschbühler (1982) pointed out some examples where ellipsis fails to disambiguate scope. The interpretation of (3:3) is that each window had both a Canadian and an American flag and therefore ‘each window’ must have scope over both clauses. Nevertheless, the following example shows that this explanation cannot be applied to similar cases:

(3:3) *A Canadian flag was hanging in front of most windows, and an American one was [] too.*

Here, the quantifier in the phrase ‘most windows’ must scope separately in each clause (cf. also Shieber et al., 1996). Other interactions of scope and ellipsis discussed in Scheffler (2005) are:

- Pronoun reference puzzles:

(3:4) *Bill likes his mother. John does [], too.* (two possible readings (John liking Bill's or his own mother):

- Dahl's puzzle (Dahl, 1973):

(3:5) *John thinks he's cute, and Bill does [], too, although his wife doesn't [].* ("John thinks John is cute, and Bill thinks Bill is cute, although his wife doesn't think Bill is cute" has been claimed to be the most natural reading.)

- Nonlocal sloppy readings:

(3:6) *Norma(1) told Beth's boyfriend to give her(1) a dime, and Judy(2) told Lois' boyfriend to [].* (correct reading [] = "give her(2) a dime")

- Multiple pronouns puzzle:

(3:7) *Bill(1) believed that he(1) loved his(1) wife, and Harry did [], too.*

[] sloppy - sloppy: "... and H. believed that Harry loved Harry's wife."

[] strict - strict: "...and H. believed that Bill loved Bill's wife."

[] sloppy - strict: "... and H. believed that Harry loved Bill's wife."

*[] strict - sloppy: "...and H. believed that Bill loved Harry's wife."

Grammatical mismatches with regard to nouns have been discussed, for instance, in Hardt (1993), Murguia (2004), Nunes and Zocca (2009) from whom the following examples are taken:

- Mismatches with regard to number (the antecedent and the omitted element have to be the same lexeme as in [3:8-10]):

(3:8) *Mickey is a mouse, but Donald and Daisy aren't [mice].*

(3:9) *Pete and Drew are children, but Chris isn't [a child].*

(3:10) *We have two people: one [person] working at the practical level of making the things happen, and one [person] who produces the information that you've seen on the website.²⁶*

- Mismatches with regard to gender:

(3:11) *Peter talked to his teacher, and Mary did too. (two possible readings)*

²⁶The words 'person' and 'people' represent the same lexeme as the unmarked plural of person is expressed by the word 'people' (cf. OED).

(3:12) ? *Brad is an actor and Angelina is [an actress] too.*²⁷

cf. Brazilian Portuguese:

(3:13) *O João é médico e a Maria também é [médica].*

the *João* is doctor-masc and the *Maria* also is doctor-fem

Maria is a doctor and João is, too

• Mismatches with regard to part of speech (verbal ellipsis with nominal antecedents):

(3:14) *John is a great laugher—when he does [laugh], his eyes crinkle up.*²⁸

(3:15) *We should suggest to her that she officially appoint us as a committee and invite faculty participation. They won't [], of course.* (example from Hardt, 1993, also discussed in Miller's paper (Miller, n.a.)) on verb phrase ellipsis with nominal antecedents)

²⁷ Graded (un)acceptability is indicated by a question mark in front of a sentence and ungrammatical or non-existing sentences are marked with an asterisk (*).

²⁸ In this example and in similar ones such as the following an auxiliary-as-proform analysis would also be possible.

- Mismatches involving negative polarity items:

(3:16) *Tom doesn't have any paper. Harry does [have some paper],
though.*

(3:17) *John didn't see anyone, but Mary did [see someone]* (example
originally from Sag, 1976b: 157).

Cases that involve reflexives are another example where syntactic
identity between the antecedent and the elided VP does not seem to be
respected:

(3:18) *John defended himself, because his lawyer couldn't [defend him].*

(3:19) *John talked about himself, but Susan didn't [talk about herself].*

In the literature on ellipsis, it has generally been observed that verbal
morphology and also that of other word classes need not always match
between conjuncts (e.g. Warner, 1986; Hardt, 1993; Lasnik, 1999a;
Lightfoot, 1999; Thoms, 2013). As Merchant (2008, 2010a, 2013) points
out, the elided material and antecedent phrase must match in voice in the
case of sluicing²⁹, gapping³⁰, stripping³¹ and fragment answers in question-

²⁹ e.g. *He invited someone, but I don't know who(m) [he invited] / Er lud jemanden ein,
aber ich weiß nicht, wen [er einlud].*

answer sequences. However, in a few cases of verbal ellipsis, voice alternations are possible: Elided verbs and their antecedents can sometimes mismatch in voice, with passive verbs being elided under apparent identity with active antecedent verbs, and vice versa. Merchant raised the question whether the head that determines voice could be external to the phrase being elided, which would mean a separation of a syntactic feature from the form on which it is morphologically realised (Merchant, 2013: 1).

Some examples of mismatches in verbal ellipsis are given in Merchant (2010b):

- Active antecedent, passive ellipsis:

(3:20) *The janitor must remove the trash whenever it is apparent that it should be [removed].*

(3:21) *'No-one can hypnotize me.' – 'Usually the people who are certain they can't be [hypnotized] are the easiest to do it to.'*³²

(3:22) *... there was really no one at the meeting who could answer the question the way it should be [answered]*³³

³⁰ cf. Ross 1970, e.g. *John likes coffee and Mary, tea. / John mag Kaffee und Mary [mag] Tee.*

³¹ Stripping / bare argument ellipsis is sometimes analysed as a subtype of gapping where only one contrast between two conjuncts is left, usually with an extra polarity element (*too, not*), cf. Repp, 2009a.

³² Example originally from the book *Simple genius* David Baldacci (2007), Grand Central: New York, p. 300.

- Passive antecedent, active ellipsis:

(3:23) *The system can be used by anyone who wants to [use it].*

(3:24) *This information could have been released by Gorbachev, but he chose not to [release it].* (example from Hardt, 1993: 37)

(3:25) *This problem was to have been looked into, but obviously nobody did [look into this problem].* (example from Kehler, 2002: 53)

- Illicit voice mismatches:

(3:26) **Someone murdered Joe, but they don't know who by [he was murdered].*

(3:27) **Joe was murdered, but they don't know who [murdered him].*

(3:28) **Jemand hat Peter ermordet, aber sie wissen nicht, von wem [er ermordet wurde].*

(3:29) **Lilies are brought by some and others roses.*

³³ 'Member comments', *Evergreen*, Newspaper of the Hyde Park Cooperative Society, Vol. 60.2, February 2007

Merchant (2001: 22) observes that gerunds can antecede sluicing in infinitival clauses, even in cases where there is a mismatch with respect to verbal morphology and the presence of infinitival complementiser ‘to’.

(3:30) *[Decorating for the holidays] is easy if you know how [to decorate for the holidays].*

Rooryck and Schoorlemmer (2013) observed that progressive ‘-ing’ can be the antecedent for a past participle in the ellipsis site (3:31) while the reverse is not impossible. They conclude that this entails that the ‘-ing’ form is morpho-syntactically more specified than the past participle.

(3:31) *Mary is leaving, and I can see that Peter already has left.*

Progressive ‘-ing’ can also function as antecedent for an infinitive:

(3:32) *Why am I saying this? Because I have to [say this].*

Ellipsis-antecedent mismatches can be expected to occur within clause complexes and do not span long distances. Ellipsis sites that are not entirely syntactically identical to the antecedent will usually be found in close proximity to the antecedent phrase. In the case of mismatches with regard to specific grammatical features, the elements of a clause are still

syntactically as symmetric as possible. In gapping as a particular type of sentence-internal verb ellipsis (cf. also Chapter 5.2.2 on gapping structures), number and person mismatches are possible (3:33) if other features of the conjuncts are symmetric.

(3:33) *Sie bauen eine Villa, ich / er [baue / baut] einen Pavillon.*

In this regard, the rules allowing verbal ellipsis resemble those that apply to nominal ellipsis where certain mismatches are acceptable if the antecedent and the omitted element are the same lexeme and are embedded in parallel structures. Mismatches with regard to the exact form of the antecedent noun and the omitted noun from the ellipsis remnant can be assumed to be more easily tolerated in German than in English. In addition to number mismatch, German can have various types of grammatical case mismatches with implicit nouns in the ellipsis sites that would be marked by a different case morpheme than the antecedent, but the elements in the ellipsis remnants (e.g. determiners, adjectives, quantifiers) are marked for case which facilitates the interpretation of the ellipsis site, e.g. in (3:34).

(3:34) *Schneller wird ein Wanderer voranschreiten, wenn ihm mehrere, ob auch kleinere Lichter (antecedent: nominative, pl.) den Weg erhellen, als wenn er sich nur eines größeren [Lichtes, genitive, sg.] bedienen kann.*³⁴

Number and person mismatches involving different forms of the omitted verb and the antecedent can be expected to occur more often in German than in English, for instance in gapping environments due to the higher number of different forms of German verbs (cf. [3:33] vs. [3:35]).

(3:35) *Why does he have everything, and they [have] nothing?*³⁵

Tense mismatch is rare in gapping as tense seems to be more relevant to the acceptability of gapping examples than other agreement mismatches. In the following case of ‘stripping’, a subtype of gapping, tense mismatch in the ellipsis is possible as only the lexical verb is left out that is preceded by ‘will’ as a remnant structure clearly indicating how to complete the ellipsis site.

(3:36) *John slept and Mary will [sleep] too.*

³⁴ example from *Beiträge zur Geschichte der Philosophie des Mittelalters*, Vol. 30, Ch. 5, p. 310, <https://books.google.de/books?id=LzYTAQAAMAAJ&pg=PA310> [last checked 08/05/2016]

³⁵ example from the book *In Black and White: An Interpretation of the South* (2008), Lily Hardy Hammond, p. 39, <https://books.google.de/books?id=gDsl4bOzzPMC&pg=PA39> [last checked 08/05/2016]

Similar cases of sentence-internal lexical verb ellipsis are (3:37a, 3:38-39). At least one modal or auxiliary verb as a finite verb is grammatically necessary as remnant in English. These verbs disambiguate the ellipsis site and specify how it should be completed (e.g. ‘*she could* [+ infinitive]’, ‘*she shouldn’t have* [+participle]’, ‘*she has* [+ *past participle*]’). The possibilities in German for lexical verb ellipsis after modal or auxiliary verbs are restricted and in German lexical verbs more often tend to be repeated within the same sentence (cf. also Chapter 7 on the relationship between ellipsis and lexical cohesion).

(3:37) a) *Leila took over every task she could [] and many that she shouldn’t have [].*

b) *Leila übernahm jede Aufgabe, die sie übernehmen konnte, und viele, die sie nicht hätte übernehmen sollen.*³⁶

(3:38) *It may come as something of a disappointment to leave your motel room for the restaurant over the way, order steaks, and then find you can order neither beer nor wine (though on your travels you could [] yesterday and will [] tomorrow).*³⁷

³⁶ Example from the novel *Purity* (2015) by Jonathan Franzen, p.187 and its German translation *Unschuld*, p. 192, <https://books.google.de/books?id=cT94BwAAQBAJ&pg=PA187> & <https://books.google.de/books?id=mirzCAAQBAJ&pg=PT192> [last checked 28/01/2016]

³⁷ Example from *Living and Working in America* 2004, Steve Mills, <https://books.google.de/books?id=1445av-V5TsC&pg=PA56> [last checked 28/01/2016] - This example is slightly marginal as the ellipsis site after ‘*will*’ should be completed with ‘be able to order beer and wine’ which has no syntactically matching antecedent.

(3:39) *And finally one day Greywing begins to cry, even more bitterly than she has [] so often since her mother has walked motionlessly on this lengthening journey. (EO_FICTION_005)³⁸*

A gapping remnant can include a temporal modifier and leave out the entire verb phrase (3:40). Some people would probably consider such examples less acceptable than others.

(3:40) *John can't leave today and Mary [couldn't leave] yesterday.*
(example from Frazier, 2015)

It is difficult to construct other examples where a tense mismatch between the gapped and antecedent constituent is acceptable or unambiguous (Frazier, 2015: 35). In most cases, at least the auxiliary verbs will be necessary to indicate tense contrasts.

Example (3:41) is a grammatically marginal case of gapping where the conjuncts are not exactly symmetric. The noun '*Schuhe*' in the plural form is followed by the verb '*sind*' ('are') and a singular noun while the second conjunct has only singular nouns that should be connected by '*ist*' (is).

(3:41) *? Schuhe sind unsere Leidenschaft und Qualität [] unser Versprechen.³⁹*

³⁸ ID of corpus text in the GECCo corpus

In examples (3:42-43) taken from the German translation of Salman Rushdie's *Two Years Eight Months and Twenty-Eight Nights* (2015: 121, 124),⁴⁰ the conjuncts are not symmetric either and therefore these sentences are grammatically awkward. There is a mismatch with regard to the function of the verb. The first part in both examples consists of a compound tense with a form of the auxiliary verb 'haben' [have] plus past participle, but the second part needs this form of 'haben' as a lexical verb in the sense of 'to possess'.⁴¹

(3:42) *Jetzt hatte er das Zeitliche gesegnet und sie [] sein Geld.*

(3:43) *Ihr Leben lang habe sie ein Auge zugedrückt, sagte sie zu ihren
Freundinnen, doch nun [] eine neue Brille.*

(3:42-43) resemble ellipsis in connection with 'zeugma' or 'syllepsis' which occurs when noun phrases or clauses share the same verb with two different senses at once (3:44-46).

³⁹ from an advertisement for a shoe store (http://www.lkz.de/cms_flipbook/marstall-magazin-eroeffnung-2015/page21.html#/20 [last checked 08/05/2016])

⁴⁰ <https://books.google.de/books?id=SjRmCQAAQBAJ> / <https://books.google.de/books?id=9h8rCQAAQBAJ&pg> [last checked 25/01/2016]

⁴¹ The English original does not use gapping and it has a form of 'have' as a lexical verb in the second part: 'Now old Elián was gone and she had his cash', p.87, / 'She had spent a lifetime turning a blind eye, she told girlfriends, but now she had new glasses' (p.89), <https://books.google.de/books?id=9h8rCQAAQBAJ&pg=PA89> [last checked 25/01/2016]

(3:44) *I am leaving for greener pastures and ten days.*⁴²

(3:45) *Er schlug die Scheibe und den Weg nach Hause ein.*

(3:46) *Sie baut ein Haus und auf Gott.*⁴³

This has been claimed to occur in the context of some gapping structures (Rooryck, 1985), but it is often better analysed as a verb linked to two coordinated noun phrases or prepositional phrases. Such constructions have been described as problematic or marginally grammatical, but they can be used for rhetorical or humoristic effects.

It should be mentioned that some cases that may be seen as cataphoric ellipses are better analysed as modifiers of the same heads or Right Node Raising (RNR) structures (cf. Chapter 5.1.8). However, Right Node Raising structures show similar possibilities for mismatches as anaphoric ellipses:

(3:47) *John WILL [sleep in her house], and Peter already HAS slept in her house* (Boškovic, 2004).

It is probable that ellipsis-antecedent mismatches will involve more complicated reconstruction operations which may lower the acceptability

⁴² example from *The Everyday Guide to Writing Wisely* by Laurie Dart p. 37, https://books.google.de/books?id=3_Uc87LXQLcC&pg=PA37 [last checked 25/01/2016]

⁴³ example taken from Hoffmann (1999)

of elliptical clauses. Several psycholinguistic studies of acceptability judgments of ellipsis have been conducted over the past years and identified a cline of acceptability for various types of antecedent - ellipsis mismatches (Arregui et al., 2006; Kim and Runner, 2009; Kim et al. 2011). Although there is a certain tension between the role which acceptability judgments play in linguistics and the level of their scientific underpinning, such experiments with more or less flawed antecedents have led to the development of a ‘recycling hypothesis’ which combines syntactic, semantic and pragmatic aspects claiming that when a syntactically-matching antecedent is not available, the listener / reader has to rely more on semantic and pragmatic aspects and reconstructs a matching antecedent using the semantic and pragmatic materials at hand (see Arregui et al., 2006, Frazier, 2013). In addition, particularly in spoken registers, parallel structures are used to mark certain focus structures and to disambiguate ellipses. In many cases, the acceptability of elliptical clauses cannot be determined without reference to the preceding linguistic discourse.

To sum up, this section has shown some interactions of scope and ellipsis that pose certain difficulties for syntactic theories. Another issue that is frequently addressed in the literature on ellipsis is the nature of identity between elided material and textual antecedents. Several examples of ellipses provide no exact syntactically-matching antecedent, but may involve mismatches, for example, in number, tense or voice etc., falling along a gradient cline of acceptability. They allow an antecedent-trigger

(i.e. the ellipsis site) that is not exactly identical to the antecedent. However, numerous examples in the literature involving mismatches have been constructed for a theoretical discussion or for testing their acceptability in psycholinguistic research. In Chapter 9.5, we will also briefly address mismatches that occur across English and German data in our corpus.

4. The classifications of ellipsis subcategories

4.1 Existing English and German ellipsis taxonomies

Several ellipsis typologies have been suggested on the basis of different underlying dimensions. Detailed taxonomies for various ellipsis subtypes in English and German have been proposed by Bühler (1990 [1934]), Quirk et al. (1985: 888ff.), Zifonun et al. (1997: 413 ff.) and Merchant (2001) among others. Ellipsis typologies often distinguish subcategories based on the assumed mechanisms of syntactic, semantic and pragmatic recoverability and interpretation of the ellipsis site as discussed in the previous chapter and frequently include ellipses as omissions and other types of fragments or non-clausal units.

Thomas (1979: 43) described ellipsis as a ‘communicative option to omit from sentences contextually available elements that are structurally required by the elements that do appear in those sentences’. He distinguished ellipsis from ‘elision’⁴⁴, which he understands as the optional omission of specific elements that may be supplied from an interlocutor’s knowledge of the language system without the aid of context. These two phenomena are both distinguished from ‘non-realization’ (Thomas 1979:49), which is the optional total absence of potential elements from a sentence, for instance as in ‘*John has been kidnapped.*’ This sentence is not grammatically incomplete although it presupposes the existence of an

⁴⁴ Elision is also sometimes defined as the loss of sounds or the omission of letters or syllables.

agent who remains linguistically unexpressed.

A distinction is drawn by Halliday and Hasan (1976: 130) between endophoric (textual) and exophoric (situational) ellipsis. Reich (2011) suggested the terms ‘*a-ellipsis*’ (antecedent-based ellipsis) and ‘*s-ellipsis*’ (situation-based ellipsis) to distinguish between omissions that have a linguistic antecedent and situational ellipsis (cf. also Schwabe [1994] on the syntax and semantics situational ellipsis). Hankamer and Sag (1976) divided anaphoric ellipses into two main categories: ‘deep’ and ‘surface’ anaphora. While surface anaphors require a suitable syntactic antecedent, deep anaphoras only require a semantic referent and can be situationally evoked.

In the German literature on ellipses with syntactic antecedents, the terms “Analepse” (in analogy with the term ‘anaphor’) or “Katalepse” for cataphoric ellipses are sometimes used (Zifonun et. al, 1997: 409ff.) to describe different types of endophoric ellipses that depend on the linguistic context.⁴⁵ Cataphoric ellipses are cases such as (4:1), but they are extremely rare in our data and they will usually not link elements across sentence boundaries.⁴⁶

⁴⁵ In my opinion, the use of the terms “Analepse” and “Katalepse” for anaphoric and cataphoric ellipsis may lead to misunderstandings as similar terms in English have different meanings. ‘Analepsis’ in English, coming from post-classical Latin *analepsis* and its etymon ancient Greek ἀνάληψις, refers to a repetition in grammar or rhetoric, the ascension or assumption into heaven, a recovery after illness, or the narration of an event at a point later than its chronological place in a piece of literature (cf. OED entry <http://www.oed.com/view/Entry/248603> [last checked 12/02/2016]). ‘Catalepsy’ and ‘katalepsis’ on the other hand, from Greek κατάληψις, are already in use with the meaning of either a disease or ‘grasping / apprehension’ in Stoic epistemology (cf. <http://www.oed.com/view/Entry/28703> [last checked 12/02/2016] and Peters 1967: 97).

⁴⁶ As already discussed in Chapter 3.3., some cases that can be seen as cataphoric ellipses (e.g. *Hans besitzt [] und sucht alte Bücher.*, discussed by Hoffmann, 1999: 75 as an

(4:1) *If you want me to [], I'll buy the tickets.*

A further selection of various subtypes of ellipsis and other fragments that have been suggested by various prominent authors is given in Table 1.

example of “Katelepse”) are better analysed as Right Node Raising structures in the context of co-ordination.

Type of ellipsis or fragment	- (Selection of) Authors who have addressed this type of ellipsis or fragment - Features	Examples
Minor clause types	Huddleston and Pullum (2002: 944ff.) main clause constructions that do not belong to major clause types	see below under a) – e)
a) optatives		<i>Would that he were still alive!</i>
b) clauses with the subordinate form	cf. also <i>Duden Grammar</i> (2009)	<i>That it should have come to this!</i> <i>To think that he was once the most powerful man in the land!</i>
c) conditional fragments		<i>Ob Anna noch anruft?</i> <i>If it isn't my old friend!</i> <i>Supposing, something happens to part us, June?</i>
d) verbless directives; Handlungsellipsen, verbless requests;	cf. also Klein (1985); Quirk et al. (1985: 842f.)	<i>Out of my way!</i> <i>Two coffees, please.</i> <i>Cigarette?</i>
e) parallel structures		<i>The sooner, the better.</i> <i>More haste, less speed.</i>

<p>gapped coordination / coordination reduction / Koordinations-ellipsen / gapping</p>	<p>e.g. Ross (1970); Huddleston and Pullum (2002: 1337ff.); Klein (1985); van Craenenbroeck (2004, 2010)</p> <p>All forms of gapping are characterised by omission of the posterior member of a pair of lemma-identical verbs. The position of this verb is often medial. Such pairs are contrastive and resemble answers to implicit multiple wh-question (e.g. <i>Who reads what?</i>).</p>	<p><i>The PM arrived at 6 and the Queen an hour later.</i></p> <p><i>His father wanted him to marry Sue, but his mother Louise.</i></p> <p><i>Max hadn't finished his assignment, nor Jill hers.</i></p> <p><i>His criticisms of Kim were inaccurate and of Pat irrelevant.</i></p>
<p>clausal comparative deletion / ellipses in comparatives</p>	<p>Huddleston and Pullum (2002: 1097ff.), Lechner (2004)</p>	<p><i>Jill can run faster than Ed.</i></p> <p><i>She went to more countries than I (vs. than me) [].</i></p> <p><i>Jones published more papers than Smith expected [].</i></p>
<p>register specific ellipses / text type specific ellipses</p>	<p>Klein (1985); Quirk et al. (1985: 838ff.) use the term 'irregular sentences' for certain incomplete / irregular structures that occur in specific text types, e.g. block language in headlines or advertisements, personal letters, diaries, abbreviated sentences in instructional writing, informal conversation, abbreviated sentences in broadcast commentaries</p>	<p><i>Germany's AAA rating under threat.</i></p> <p><i>Sorry about Jane.</i></p> <p><i>Having wonderful time.</i></p> <p><i>Want another cup?</i></p> <p><i>Heat to boiling point.</i></p> <p><i>Contains whitener and brighteners.</i></p> <p><i>Don't know where they are.</i></p> <p><i>The first lap is over. Five more to come.</i></p>

<p>adjacency ellipses / fragment answers / elliptical sentences in dialogue, echo utterances</p>	<p>e.g. Klein (1985); Quirk et al. (1985), <i>Duden Grammar</i> (2009: § 1396)</p> <p>ellipsis in the second component of an adjacency pair such as question-answer pair, corrections, confirmations, echo utterances, adding more information</p>	<p><i>How many students were there? – Three.</i></p> <p><i>This cake is for Mary – No, for John.</i></p> <p><i>Peter hat Glück im Spiel - Ja, im Spiel. In der Liebe auch.</i></p> <p><i>The Browns are emigrating. – Emigrating? (yes-no question)</i></p> <p><i>I'll pay for it? – You'll what? (wh-echo)</i></p> <p><i>Wann wird der Hundertmeterlauf beginnen? – Wie bitte? – Wann der Hundertmeterlauf beginnen wird?</i></p>
<p>expressive exclamations</p>	<p>Klein (1985); Quirk et al. (1972: 7.84ff.)</p>	<p><i>Great!</i></p> <p><i>Charming couple!</i></p> <p><i>The cake! (self-addressed → I should have taken the cake out of the oven!)</i></p>
<p>elliptical formulaic expressions</p>	<p>Klein (1985)</p>	<p>No idea.</p> <p>Schon möglich.</p>
<p>lexicalised ellipsis / ellipsis of complement of lexical verbs and adjectives</p>	<p>Klein (1985), Ágel (1991), Huddleston and Pullum (2002: 1527ff.)</p> <p>conventionalised, context independent 'ellipsis'</p>	<p><i>Wer gibt (Karten)?, Er hat zwei Jahre (im Gefängnis) gegessen.</i></p> <p><i>It's your turn to deal (the cards)</i></p>
	<p>Klein (1985)</p>	

ellipsis due to processing difficulties	production of an incomplete sentences if speaker has difficulties in finishing a sentence or realises that his utterance has become unnecessary or inappropriate	
ellipses due to incomplete linguistic development	Klein (1985) ellipsis produced during language acquisition processes or because of language disorders	
sluicing	Ross (1969), Merchant (various) ellipsis of an interrogative clause to the sole exclusion of the wh-word	<i>Max went to the store, but Oscar wondered why.</i> <i>Somebody messed up the printer. Guess who.</i> (a special case is multiple sluicing where the second remnant wh-phrase is a prepositional phrase -> <i>I know that in each instance one of the girls got something for one of the boys. But [which] [for which]?)</i>)
reduced conditionals	Schwarz (1998) The remnant in a reduced conditional bears the same case as it would in a nonelliptical conditional; the 'dann'-clause contains an elided clausal structure can be considered a subtype of coordination reduction / gapping	<i>Wenn der Hans wen besucht, dann [] {*der / den} Peter.</i> <i>if the NOM Hans someone ACC visits, then *the NOM / the ACC Peter</i> <i>'If Hans visits someone, then it's Peter.'</i> [German]
specificational pseudoclefts	den Dikken (2005) can be considered a subtype of	<i>What they didn't buy was [they didn't buy] any wine.</i>

	coordination reduction / gapping, but clefts and pseudoclefts are no typical cases of ellipsis	
spading	<p>van Craenenbroeck (2004: 13ff.)</p> <p>sluicing plus a demonstrative in non-insular Germanic: in a number of Dutch dialects, Frisian, Norwegian, French and some German dialects a sluiced wh-phrase can be followed by a demonstrative pronoun</p>	<p><i>Jef eid iemand gezien, mo ik weet nie wou da.</i></p> <p><i>(Jeff has someone seen but I know not who that DEM</i></p> <p><i>'Jeff saw someone, but I don't know who.'</i></p> <p><i>[Wambeek Dutch])</i></p>
non-sentences	<p>Quirk et al. (1985: 849ff.)</p> <p>often NPs, in speech (with specific intonation patterns or gestures):</p> <p>commands, requests, exclamations, invitations, inquiries, introductions (vocatives), formulae (greetings, thanks, reaction signals (assent, agreement/denial, disagreement), toasts, warnings, apologies, expressions of anger, expletives...), interjections</p>	<p><i>The clothes she wears!</i></p> <p><i>You and your statistics!</i></p> <p><i>New hat?</i></p> <p><i>Your book or your mothers?</i></p> <p><i>Joan, my suster – John, a good friend of mine.</i></p> <p><i>Good morning!</i></p> <p><i>Really?</i></p> <p><i>My mistake.</i></p>
strict ellipsis	<p>Quirk et al. (1985: 888ff.)</p> <p>grammatical omission (in contrast to semantic omission)</p> <p>- criteria: a) precisely recoverable, b) defective, c) insertion of missing expression results in grammatical sentence with same meaning, d) recoverable from the neighbouring text</p>	<p><i>I am happy if you are [].</i></p> <p><i>She can sing better than I can [].</i></p> <p><i>If you want me to [] I'll lend you my pen.</i></p> <p><i>We try, whenever we can [], to leave the window</i></p>

	(rather than from structural or situational context), e) missing expression is exact copy of antecedent (strict ellipsis is applicable mainly to coordination)	<i>open.</i> <i>Those who prefer (to) [] can stay indoors.</i> <i>Don't ask me why [], but the stone has been moved.</i>
standard / general ellipsis	Quirk et al. (1985: 888ff.) e) is not necessary (see above)	<i>She sings better than I can [*sings].</i>
quasi-ellipsis (on the borderline between ellipsis and substitution)	Quirk et al. (1985: 888ff.) full form cannot be recovered without changing objective pronoun	<i>She works harder than him (= he does).</i> <i>She understands the problem better than him (=he does).</i> <i>possessive pronouns: mine, yours, theirs etc. (apart from 'his') / do-support construction -> do as a stranded operator</i>
situational ellipsis / Kontextellipse	Quirk et al. (1985: 888ff.); Klein (1985) d) and e) not necessary (see above)	<i>(I am) Glad to see you.</i>
ellipsis without precise recoverability	Quirk et al. (1985: 888ff.) a) not necessary (see above)	<i>(Since he was / Being) Angry, he stalked out.</i>
structural ellipsis (sometimes not precisely recoverable) /	Quirk et al. (1985: 888ff.) structurally recoverable (e.g. relative pronoun-dropping, that-deletion, asyndeton [omission of conjunctions])	<i>I believe (that) you are wrong.</i> <i>The man (that/who/whom) I saw was half asleep.</i>
non-finite clauses	Quirk et al. (1985: 888ff.)	<i>Houses (that/which are)</i>

	similar to ‘structural’ ellipsis	<i>owned by Mr. Smith...</i>
semantic implication	Quirk et al. (1985: 888ff.) end-point of ellipsis gradient, is more fittingly classified not as ellipsis	<i>The door opened and (then/after that/...) Mary entered (although the concept of sequence in time is understood in the sentence, there is no reasonable way of choosing between different adverbials).</i>
special types of ellipsis	Quirk et al. (1985: 888ff.); Klein (1985) in non-finite and verbless clauses, in coordination, in comparative clauses, in appended clauses	
pronoun / topic dropping	various	[] <i>Wish you were here.</i> [] <i>Ist klar.</i>
pseudogapping	Merchant (various) variant of VP ellipsis, in which some VP internal constituent is not affected by ellipsis	<i>John could pull you out of the plane, like he did [] his brother.</i> <i>Alicia ate a piece of the cake, and Beatrix did [] a slice of cheese.</i>
null complement anaphora	Hankamer and Sag (1976) the verbal predicates that can license null complement anaphora form a limited set (e.g. know, approve, refuse, decide)	<i>Mary didn't notice.</i> <i>I really tried.</i> <i>They offered two ways to spend the day, but I couldn't decide.</i>
nominal ellipsis / noun ellipsis	Halliday and Hasan (1976: 147ff.), Günther (2013)	cf. Chapter 5.1

	ellipsis within the nominal group	
verbal ellipsis / VP-ellipsis	Halliday and Hasan (1976: 167), Hardt (1993) ellipsis within the verbal group presupposing either the lexical verb or the operator, the deletion of a verb phrase in a sentence	cf. Chapter 5.2
clausal ellipsis	Halliday and Hasan (1976: 196ff.)	cf. Chapter 5.3

Table 1: Ellipsis subtypes according to different authors, source: own table, based on literature reviewed

One of the most prominent classification systems among those given in the list above is the one suggested by Quirk et al. who assumed a continuum of ellipsis and proposed a definition of ‘strict’ ellipsis and less ‘strict’ forms of ellipsis (1985: 888ff.). The criteria for strict ellipsis are:

- The elided words are precisely recoverable.
- The elliptic construction is grammatically “defective”.
- The insertion of the missing words results in a grammatical sentence with the same meaning as the original sentence.
- The missing words are textually recoverable.
- The missing words are present in the text in exactly the same form.

A strict ellipsis is therefore defined as an incomplete, grammatically defective structure that can be understood from the surrounding text. It involves verbatim recoverability of the missing elements, i.e. the omitted words are precisely recoverable. The elements left out have to be present in the text in exactly the same form while the insertion of the missing words results in a grammatical sentence with the same meaning as the original sentence. Less 'strict' forms of ellipsis that meet only some of these criteria fall under 'standard ellipsis', 'situational ellipsis', 'structural ellipsis', 'weak ellipses' or 'quasi-ellipsis'. Quirk et al. rank their definition criteria assuming that some are more important than others and describe ellipsis as a gradual phenomenon, which could lead us to think of ellipsis as a prototype continuum with core and peripheral members.

Quirk and Greenbaum (1990: 256) additionally distinguish three categories of ellipsis depending on the place where the ellipsis occurs within the construction. They divide omissions into ellipses of initial, medial or final elements. This has sometimes led to a rather simplified description of textual ellipses in pedagogical grammar books where only initial, medial and final ellipsis are given as categories regardless of other aspects (e.g. Downing, 2014: 225). Sanders (1977: 255) had already presented a similar typology of ellipsis constructions with a specific focus on ellipsis in coordinated structures where he analysed ellipsis depending on their position in a coordinated sentence. Sanders identified six different

positions in which ellipsis can occur in the structure [A B C & D E F], cf. also Haspelmath (2007: 44f.).⁴⁷

Another oft-cited categorical ellipsis typology among the ones listed in Table 1 is the one proposed by Klein (1981, 1984, 1985, building on Bühler, 1990 [1934]) who, with a particular focus on German, defined subcategories of regular ellipsis (“regelhafte Ellipsen”) where contextual information may be taken from preceding or following utterances, from the perceivable situation or from factual knowledge. Regular ellipsis means that these types of ellipses are resolved by applying certain grammatical rules. Nevertheless, not all of the following categories, suggested by Klein, follow consistent rules for their resolution:

- text-type specific ellipses (“Textsortenellipsen”),
- ellipses as orders to perform actions (“Handlungsellipsen”),
- expressive exclamations (“expressive Ausrufe”),
- elliptical formulaic expressions (“elliptische Formeln”),
- lexicalised or conventionalised ellipses (“lexikalische Ellipse”),⁴⁸
- coordinate ellipses (“Koordinationsellipsen”),
- adjacency ellipses (“Adjazenzellipse”),
- ellipses due to processing difficulties (“verarbeitungsbedingte

⁴⁷ Sanders (1977:255) classified ellipsis types according to their position in an abstract pattern “ABC & DEF”. This leaves us with six logically possible types of ellipsis: []BC & DEF: A-ellipsis (‘initial catalipsis’) / A[]C & DEF: B-ellipsis (‘medial catalipsis’) / AB[] & DEF: C-ellipsis (‘final catalipsis’) / ABC & []EF: D-ellipsis (‘initial analipsis’) / ABC & D[]F: E-ellipsis (‘medial analipsis’) / ABC & DE[]: F-ellipsis (‘final analipsis’).

⁴⁸ not to be confused with ‘lexical ellipsis’, a term that is used by Halliday & Hasan (1976) for the omission of the lexical verb

Ellipsen”)

- ellipses due to incomplete linguistic development
(“entwicklungsbedingte Ellipsen”)
- other types of ellipses

Zifonun et al., 1997: 409ff. suggested a further category referring back to Bühler’s terminology: “empraktische Ellipse” (‘empractic ellipsis’), which we might also call ‘pragmatic ellipsis’. ‘Empraxis’ means that language is intertwined with non-linguistic social activities. An utterance is ‘empractic’ if it is embedded in a series of activities (Auer, 1988: 268). This type of ellipsis is in a way similar to the above-mentioned text-type-specific ellipses and ellipses as orders to perform actions. The interlocutors need to have pragmatic knowledge about the discourse context and social actions. To understand an utterance such as *‘Ready, steady, go!’*, it is necessary to know the actions that take place on a racetrack.

Such existing typological schemes do not always place ellipsis subtypes on an equal footing and turn out to have grey areas or overlaps between their categories, which makes them highly non-operational for corpus linguistic analyses. Different dimensions of comparison, for instance aspects with regard to form, function, context and recoverability, are reflected in the classification schemes so that those categories do not have the same level of abstraction in the taxonomy. Various aspects also mix in Klein’s ellipsis classifications and tend to overlap. In his typology, ellipsis categories are defined either on the basis of their form, their syntactic or

textual context, their level of conventionalisation or the reason or intention behind their use.

As can be seen in the list given in Table 1, on the one hand, subcategories in ellipsis typologies are frequently defined by the structure of the element that is left – the ellipsis remnant – e.g. in expressive exclamations or stranded auxiliaries. On the other hand, other subcategories are defined with regard to the element that is left out (e.g. lexical verb ellipsis) or with regard to the assumed complete constituent or structure in which something has been omitted (e.g. clausal ellipsis). Additional criteria that have been applied for determining subcategories are contexts in which certain ellipses are used (e.g. situational ellipsis, empractic ellipsis, elliptical sentences in dialogue) or functions they fulfil (fragment answers, orders to perform actions etc.). It is important to bear in mind that formal, functional and contextual aspects sometimes mix to a certain degree in existing ellipsis typologies, which contributes to the complexity of the discussion about ellipsis categories.⁴⁹

⁴⁹ Here it may help to draw an analogy to the classification principles for organisms and species in the field of biology as no other science has developed a more nuanced and sophisticated understanding of the epistemology of classification. In the development of classification and identification schemes in biology, it was recognised that non-morphological and non-structural characters such as relation to man ('cultivated' versus 'wild'), seasonality ('deciduous' or 'evergreen') or ecological affinities and habitat (forest, marsh, etc.) were less apt to produce useful classifications than strictly structural characteristics. Consequently, the use of structural characteristics has become the dominant aspect in taxonomy since the sixteenth century (cf. chapter "Macrotaxonomy, the science of classifying" in Mayr, 1982). Similarly, in our case, structural features of ellipsis as a grammatical concept should be the dominating aspect in the classification. Aspects such as contexts or text-types in which these patterns occur or acceptability ('proper' vs. 'improper' use) play a subordinate role and are considered after the structural features have been taken into account.

Cross-linguistic differences and similarities in ellipsis have been addressed in a few studies with regard to certain elliptical constructions. Particularly, research projects in the framework of transformational grammar have contributed to this field of linguistics, e.g. Aelbrecht (2010) on ellipses in Dutch and English. Goldberg (2005) gives a cross-linguistic analysis on verb-stranding VP ellipsis and Merchant and Simpson (2012) provide a cross-linguistic perspective on sluicing. Other specific constructions in various languages have been described in comparison to English, for instance, by van Craenenbroeck (2004) for Dutch, van Craenenbroeck and Lipták (2006) for Hungarian, Hinds (1982) and Makino (1993) for Japanese, Shopen and Świczkowski (1976), Duczmal (1985) and Juzwa (2006) for Polish, Zagona (1988a/b) for Spanish and İnce (2009) for Turkish.

This section discussed various existing typologies and accounts of ellipsis from different theoretical frameworks. Among these classification systems, there are only very few studies examining ellipsis use in textual contexts, and the description of ellipses types is as yet hardly related to data of their distribution in corpus data. Many approaches involve a micro-level analysis of invented examples of highly typical, simplified or even marginal and rare cases of ellipses and many of the above-mentioned publications focus in detail on the syntactic analysis of certain cases.

As the current study is embedded in a corpus-linguistic project on textual cohesion that aims to provide information about textual contrasts in cohesion between English and German and between written and spoken

registers, it is necessary to develop a cross-linguistically applicable and empirically supportable categorisation and annotation scheme. In such a scheme, we will focus on the potential cohesive function of ellipses in discourse and sort out other types of non-cohesive sentence fragments, non-clausal units and other omission phenomena that may look similar but actually need different analysis. The next section will explore Halliday and Hasan's systemic functional conception of ellipses as cohesive devices which is taken as the main basis for the development of the annotation scheme.

4.2 Ellipsis as a cohesive device in the systemic functional approach

As the previous sections demonstrated, the topic of ellipsis offers a potentially limitless body of material from the literature. This study will concentrate particularly on ellipsis used as a cohesive device. The myriad of fragmentary constructions that have been studied under the notion of ‘ellipsis’ in the past usually do not have a focus on textual cohesion and there is no comprehensive study so far that addresses ellipsis as a cohesive device with regard to the English-German language pair. A classification of ellipsis used as a cohesive device should be both theoretically consistent and easily applicable to real language data from corpora. Nevertheless, forming categories and organising them into coherent systems so that the classification of all examples occurring in the corpus is consistent with theoretical assumptions is a challenge. It is necessary to subsume the multitude of categories suggested in the literature under more general and abstract categories to cover the variety of different omission possibilities with the potential to be used as cohesive devices in English and German corpus texts from a wide spectrum of communication scenarios on written-spoken and formal-informal continua.

Halliday and Hasan’s (1976) distinction between nominal, verbal and clausal ellipses as omissions that function as potentially cohesive devices is taken as a starting point for the development of the theoretical framework for this contrastive study, ensuring the consistent use of SFL terminology within the GECCo project. These three ellipsis subtypes occur

both in written and spoken language in English and German with different register-specific frequencies. Cohesive ellipsis can be expected to be less frequent than various other types of cohesive devices. We can compare English and German nominal, verbal and clausal ellipses based on the similarities between these languages with regard to the type of structures that can contain omissions. Normally, when categories are equated across languages, surface forms are compared and different languages usually do use exactly the same morpho-syntactic forms to instantiate a category. With ellipsis, we cannot compare surface forms of overt structures. We compare the patterns of contexts in which ellipses occur and we derive the internal structure of ellipses from abstract, underlying structures. The categories of nominal, verbal and clausal ellipsis that had first been described for English as a particular language are broad or abstract enough to instantiate cross-linguistic categories for English and German and to serve as a *tertium comparationis*. Thus, these ellipsis types can be regarded as cross-linguistically valid categories or comparative concepts for the English-German language pair that we can use to describe the different instantiations of ellipsis in these languages (cf. also Haspelmath, 2010a on comparative concepts and descriptive categories in cross-linguistic studies).

In the 1970s, Halliday and Hasan dedicated a chapter of their book *Cohesion in English* (1976) to ellipsis, emphasising the possibility for ellipsis-antecedent relations to go beyond single sentences and thereby link parts of a text and contribute to its cohesiveness. The term ‘cohesion’,

popularised by Halliday and Hasan, refers to the text-internal relationship of linguistic elements that are overtly linked via lexical and grammatical devices across text segments in a coherent text. Cohesive devices are usually described as surface structure phenomena (cf. de Beaugrande and Dressler [1981: 3] and Baker [1992: 180]). Halliday and Hasan did not distinguish between different terms for the concepts of textual cohesion and coherence, but this distinction has been emphasised in other works on discourse relations (Beaugrande and Dressler, 1981). The recognition of textual coherence involves text- and reader-based features and refers to the logical flow of interrelated ideas, thus establishing a mental textual world. Relations of cohesion can be regarded as indicators of meaning relations in a text and hence contribute to its overall coherence.

According to Halliday and Hasan, ‘cohesion occurs where the interpretation of some element in the discourse is dependent on that of another’ (1976: 4). They state that ‘[...] cohesive relations are the same whether their elements are within the same sentence or not. [...]’ (ibid.: 9). Halliday and Hasan clearly state that cohesion ‘is a relation to which the sentence, or any other form of grammatical structure, is [...] irrelevant’ (1976: 9) and that cohesive ties between sentences stand out more clearly because they are the ONLY source of texture whereas within the sentence there are structural relations as well. Therefore the cohesive effect is less pronounced within the sentence (ibid.: 9). The examples they give are of cohesion across sentence boundaries assuming that in those cases the effect is more striking.

One way to hold a text together is by deleting certain textual elements. Certain types of ellipsis are an important grammatical means of achieving cohesion in written and spoken discourse. Halliday and Hasan claim that language does not function in isolation, but in actual situations of use and in texts (ibid. 142), and they associate ellipsis with textual metafunction in the systemic functional model. Research on ellipses in other approaches than the systemic functional one have focused less on the textual aspects of ellipses, although a few scholars from other approaches have also pointed out the relationship between ellipses and discourse coherence and that the acceptability of elliptical clauses cannot be determined without reference to the surrounding linguistic discourse (Kehler 1995, 2000, 2002; Frazier and Clifton, 2006).

Halliday and Hasan's taxonomy of five major categories of cohesive devices, namely reference, conjunction, lexical cohesion, substitution and ellipsis, has become a widely used and internationally recognised typology and their book has been referred to as 'the standard reference on cohesion' (Stubbs, 2001: 310). In Halliday and Hasan's scheme, a high proportion of the elements in a text are markers of textuality and continuity of ideas. Semantically related content words as well as function words such as pronouns, definite articles, grammatical substitutes and connectives create cohesive links. In this scheme, ellipsis and substitution have been characterised as the strictest cohesive relation – a purely textual relation with no other function than that of cohering one piece of text to another (Halliday and Hasan, 1976: 226).

Various slightly different classifications of cohesive devices have been suggested in subsequent literature. De Beaugrande's classification (1980) and that from de Beaugrande and Dressler (1981), for instance, comprise more main categories of cohesive devices than Halliday and Hasan's: recurrences, parallelism, paraphrases, proforms, ellipsis, junction, tense, aspect, functional sentence perspective and intonation. Nevertheless, they follow Halliday and Hasan closely with regard to their understanding of ellipsis as a cohesive device and when ellipsis has been addressed in later studies in the context of textual cohesion and coherence, Halliday and Hasan's classification of ellipsis and its subtypes as a means of grammatical cohesion was usually taken as it is and applied to the description of prototypical elliptical constructions relating passages in a text. I observed that some linguistics papers and pedagogical textbooks suffer from the lack of a coherent view of ellipsis as a cohesive device. Textbooks on cohesion and discourse analysis frequently give a rather vague or sketchy definition of ellipsis or illustrate this phenomenon with few examples, which in my opinion do not always reflect the concept of ellipsis correctly. In the book *Working with Discourse* by Martin and Rose (2013: 167), for instance, ellipsis is defined as an "implicit reference" or as "referring to participants by leaving them out". This definition is different from how ellipsis is described in most major grammar books and it is also different from Halliday and Hasan's conceptual framework on ellipsis as a cohesive device in which ellipses is the omission of elements normally

required by the grammar which can be assumed to be obvious from the context.

Martin and Rose (2013: 167) gave the following example as an illustration of cohesive ellipsis (4:2). Instead of analysing the second part of this passage as containing two cohesive ellipses of a pronoun⁵⁰ as Martin and Rose did, we favour an analysis as a sentence fragment in the form of a sentence-split (Chapter 6.2).

(4:2) *Suddenly, at strange times, they would become restless. Abruptly mutter the feared word 'trip' and drive off.*

Several other recent references on discourse analysis and textual cohesion restrict themselves to one or two paragraphs on ellipsis and to a very basic, mundane definition in which they summarise and reduce Halliday and Hasan's description of ellipsis by pointing to the general and slightly vague statement by Halliday and Hasan that ellipsis is something which is left unsaid but nevertheless understood (Flowerdew, 2012: 37).

The present study is designed to contribute to the research activities and results of a larger corpus-linguistic project on textual cohesion that aims to provide information about textual contrasts in cohesion between English and German and between written and spoken registers (GECCo). The principles of referencing, substitution and conjunction have already been addressed in great detail in the first phase of the project while lexical

⁵⁰ It is not only the pronoun that is not repeated here, but the finite verb as well.

cohesion and ellipsis are currently under investigation in the second project phase. The GECCo project's overall conceptual framework has been developed after a review of relevant literature on textual cohesion and is largely based on Hallidayan systemic functional linguistics adapted to the needs of a cross-linguistic study. Since the categorization of cohesive devices as described by Halliday and Hasan is now widely used and since the GECCo project as a whole takes Halliday and Hasan's conceptual framework as a starting point, it makes sense to also try to use Halliday and Hasan's notion of ellipsis and its further division into nominal, verbal and clausal ellipses as the basis for the development of a linguistic annotation scheme. The aim of developing annotation guidelines for all types of cohesive devices in the GECCo corpus was to utilise the categories suggested by Halliday and Hasan, to reflect upon their cross-linguistic validity and to apply them to our corpus data to empirically measure and track cohesive devices in texts of different registers. The combination of SFL as a theoretical framework with a quantitative analysis of corpus data leads to insights about the lexico-grammatical properties of English and German and about text-type-specific usage preferences of cohesive devices that would otherwise be missed. There are strengths and limitations of using this conceptual model in a very strict way that will be addressed in this chapter. Some clarifications are necessary when working with the tripartite ellipsis typology of nominal, verbal and clausal ellipsis, for which Halliday and Hasan also give various subcategories.

The main advantage of utilizing such relatively broad main categories of ellipsis is that this provides a suitable framework for a cross-linguistic analysis of ellipses with possible textual antecedents as both English and German have noun phrases, verb phrases and clauses where certain elements can be omitted that are deducible from the co-text. However, we run into several practical difficulties if we adhere strictly to the details of Halliday and Hasan's analysis, especially when we apply the classification to real discourse data from different text types and various communication scenarios in English and German and not only to idealised or prototypical isolated utterance pairs. Some of the 'grey areas' that may arise in applying Halliday and Hasan's terminology to the corpus data used for this study have already been addressed briefly in Menzel (2016b). We can slightly refine and evolve the concepts to make sure that we can apply them to English-German corpus data without having to rely on primarily gradient or prototype-based generalisations. Halliday and Hasan and later works from a systemic functional linguistic perspective remained somewhat vague on the theoretical basis for their conceptualization of ellipsis. Studies focussing on ellipses as textually cohesive ties are relatively rare and conceptually not always very well defined.

To my knowledge, Clarke's dissertation (2012) is the only substantial study looking at ellipsis from an SFL perspective and providing a critical examination of the ellipsis description suggested by Halliday and Hasan (1976). Clarke confirms that, although ellipsis has sometimes been addressed in the systemic functional literature, in the majority of these

instances, ellipsis is only covered briefly or raised as an aside in the discussion of some other topic (Clarke, 2012: 66). I agree with Clarke's statement that the concept of ellipsis has to be confined to reasonable limits, rather equating it with simply anything understood but not said 'in which case it would be so broad to be of little descriptive worth' (ibid.: 63). Clarke's main objective however was not to make quantitative statements about ellipsis and therefore he defined the different types of ellipsis along a continuum of prototypicality according to aspects such as their precise textual recoverability.

At any rate, when discussing ellipsis in the SFL framework, we cannot ignore the vast body of literature on ellipsis from other well-established approaches as this would lead to a superficial theoretical understanding of ellipsis. If we want to annotate and analyse cases of cohesive ellipsis in a corpus in order to show how this phenomenon manifests itself in written and spoken discourse and if we want to describe English-German contrasts in the use of ellipsis as a cohesive device, a conceptual clarification and elaboration of Halliday and Hasan's work on ellipses is necessary. Otherwise it is impossible to operationalise and quantify this concept and we would come to the same conclusion as Halliday and Hasan: 'Being able to give a theoretical definition in these terms does not mean, however, that for every instance [...] we can always recognise whether it is elliptical or not...' (Halliday and Hasan, 1976: 168). The classifications used for a corpus annotation should not overlap or involve gradual categories or grey areas. The aim is to place all cases found in the corpus clearly in only one

category in order to provide the basis for a meaningful quantitative analysis.

Halliday and Hasan were not the first to point out the possibility of using ellipsis for the connection of sentences, but they elaborated, for example, on what had been briefly sketched in grammar books from that period such as the *Grammar of Contemporary English* by Quirk et al. (1972) and the *University Grammar of English* by Quirk and Greenbaum (1973a). Quirk and Greenbaum (1973a: 251-253, 305-308) had covered some material on ellipsis, providing a few remarks on ellipsis in coordination and ellipsis not dependent on linguistic context in contrast to ellipsis as a means of sentence connection, for which they mainly gave examples of omissions in the second parts of sentence pairs such as question-answer sequences. Listing discourse reference, connectors, substitution by proforms, ellipses and lexical connections, e.g. the use of repetition or synonyms, as the main categories of means to connect sentences, Quirk and Greenbaum (ibid. Chapter 10) had paved the way for Halliday and Hasan's work on cohesive devices that expanded on these five categories in more detail. They put the concept of sentence connectors into a larger perspective, discussing them as signals for the connections between passages of text and not only sequences of sentences.

From the 1970s onwards, gradually more specific publications on the subject of endophoric ellipsis with textual antecedents appeared, but Halliday and Hasan did not explicitly include any reference with a particular focus on ellipsis in their work on cohesion apart from an article

on elliptical sentences in American English by Gunter (1963). Although the ellipsis discussion had been going on for centuries as explained in the previous chapters and despite some influential publications on ellipsis and its subcategories from the 1970s, e.g. Ross (1969, 1970), Jackendoff (1971), Hankamer (1973) and Kuno (1976) on gapping and sluicing as ellipsis phenomena, and a very detailed dissertation by Sag (1976) on the syntax of verb phrase deletion, Halliday and Hasan did not explicitly use the terminology from the scholarly publications from that time or from earlier specific works on ellipses and rather referred to the everyday usage of the word ellipsis. For instance, when Halliday and Hasan discussed the following example: *‘Joan brought some carnations, and Catherine some sweet peas.’* (1976: 143), they did not use the term ‘gapping’ and described it broadly as ellipsis within the sentence in the context of coordination. In fact, at the time when Halliday and Hasan’s book appeared, most publications on ellipses from that period were mainly interested in deletions in coordinate structures (e.g. Hankamer’s dissertation from 1971), or in minor and fragmentary sentences in the domain of colloquial, spoken language (Bowman, 1966), which may have been the reason for Halliday and Hasan not to consider them. The earliest SFL publications by Hasan are from the 1960s and pre-date much of the discussion and the literature on ellipses that developed in the 1970s.

The starting point of the discussion of ellipsis for Halliday and Hasan was ‘the familiar notion that it is “something left unsaid”’ and the claim that ellipsis is ‘something understood’ (ibid.: 142). This non-technical,

broad view on ellipsis simply as “something” which is implicit or unsaid initially may seem to lead to a situation where ellipsis is a relatively frequent phenomenon. If we narrow it down to a precise definition of what is meant by *cohesive* ellipsis, we have to exclude many things that are implicit or unsaid. It becomes clear that Halliday and Hasan understand ellipsis primarily as incomplete structures and as a syntactic phenomenon.⁵¹ They described ellipses as a structural mechanism distinct, for instance, from reference which is a relation between meanings. Although, like all cohesive devices, ellipsis also contributes to the semantic structure of the discourse, it sets up a relationship that is lexico-grammatical – a relationship in the wording (Halliday, 1985: 296, Halliday and Matthiessen, 2013: 635) – where the omitted element is present in an underlying structure, similarly to interpretations by Sag (1976) and Williams (1977), inter alia. ‘Ellipsis occurs when something that is structurally necessary is left unsaid, there is a sense of incompleteness associated with it...’ (Halliday and Hasan, 1976: 144). “Something” is not realised in the surface structure of a sentence and exists as a gap to be

⁵¹ Cohesive ellipses are controlled by various syntactic, semantic and stylistic factors and, in contrast to other cohesive ties, represent relations not between two actual text segments, but between an actual passage and a ‘virtual or theoretical complete’ (sic!) version (cf. de Beaugrande, 1991: 252). Being aware of the complexity of the syntax-semantics interface and the impact of pragmatic and prosodic aspects and focus structure on utterance and text processing discussed in Chapter 3, I choose to follow the syntactic approach in the current stage. Although I assume that, on the level of logical representation, ellipses are complete propositions with gaps that can be understood depending on the linguistic or situational context or the cultural knowledge of the interlocutors, what is important for the classification of my ellipsis categories is the presupposed complete clausal or phrasal structure assuming that ellipsis involves unpronounced linguistic material. Cohesion, according to Halliday and Hasan (1976), is generally a semantic concept, where the interpretation of some element in the discourse is dependent on that of another, but ellipsis additionally sets up a relationship with its textual antecedent that is lexico-grammatical, where the omitted element is present in the underlying syntactic structure.

filled from ‘elsewhere’ (ibid.: 143). The content of such a gap is licensed by structural identity between the elided element or constituent and its antecedent. Halliday and Hasan were referring specifically to ‘sentences, clauses, etc. whose structure is such as to presuppose some preceding item’ (ibid.). Therefore, in ellipsis, there is some presupposition in the structure of what is to be supplied. An ellipsis itself should not be called a cohesive *relation* (Halliday and Matthiessen 2013: 635) in the same sense as other categories of cohesive devices establish relations between different sentences (e.g. conjunction) or are relations themselves (e.g. co-reference as a relation of semantic identity between a pair of textual items). A cohesive relation can be established between the ellipsis site and its antecedent. Ellipses are not restricted to specific words or word classes. An ellipsis can involve the omission of one or several words from a larger structure. The omitted words as well as the surrounding remnant structures have to be considered to understand the nature of the textual link created by an omission.

A potential overlap of categories may result from the fact that Halliday and Hasan do not keep their five subcategories of cohesive devices strictly separate from each other, and some phenomena of textual cohesion may co-occur in one textual element as a ‘double cohesive tie’ (Halliday and Hasan, 1976: 157) that is both grammatical and lexical, for instance, when omitted words have reference as well. Additionally, ellipses can be preceded by words that express personal, demonstrative and comparative reference (cf. Halliday and Hasan, 1976: 37ff.). Halliday and Hasan define

ellipsis in relation to another cohesive device, substitution, and claim that both embody a comparable relation between parts of the text. Discussing substitution and ellipsis in distinct chapters, they suggest that the break between these chapters is an unnatural one. In a way, ellipsis is similar to cases where a grammatical placeholder element such as ‘one’, ‘do’ or ‘so’ is used. Therefore, Halliday and Hasan characterised ellipsis as ‘substitution by zero’ (1976: 143), a notion that we still find in many recent publications where, by analogy with the concept of zero phonological elements and zero-morphemes, a zero element is assumed to replace a structure. Tutin et al. 2000 refer to ellipsis as empty anaphoric expressions. I understand ellipsis as an omission where a slot is left empty, which is not a subtype of substitution.

Ellipses have to be distinguished from other types of fragments and non-clausal units, other omission phenomena and non-ellipses that might look similar but actually need different analyses. Tucker (2007) discussed certain subtypes of ellipsis in the SFL framework. He examined how non-clausal units that we frequently find in spoken language as a result of conversational routines leading to a high degree of formulaicity and fixedness of the language can be accounted for in a systemic functional grammar. He did not aim to propose a dedicated grammar of speech, but claimed that a linguistic solution that simply treats all non-clausal expression as reduced forms of some underlying fully clausal expression would miss the point. I will treat such cases under a separate category of fragments, a category that will be explained in more detail in Chapter 6.

In analogy with nominal, verbal and clausal substitution, Halliday and Hasan categorised ellipsis into nominal, verbal and clausal subtypes. Nominal ellipsis is an omission within the nominal group.⁵² The head noun is omitted so that only a modifier, e.g. a numeral, remains as the remnant. The following examples from the English and German part of the GECCo corpus illustrate this ellipsis subcategory that is rather similar in English and German in this case (4:3).⁵³ These nominal ellipses link to an antecedent across sentences.

(4:3) a) *There are many reasons why Britain is good for Europe. Let me choose just four [].* (EO_ESSAY_013)

b) *Es gibt viele Gründe, warum Großbritannien gut für Europa ist. Ich möchte vier [] herausgreifen.* (GTRANS_ESSAY_013)

Verbal ellipsis is ellipsis within the verbal group (or verb phrase), the omission of a modal, auxiliary or lexical verb (4:4)

(4:4) *Brass bands had not been invented during the time of Giraldus Cambrensis. If they had been [], he certainly would have commented on them in detail.* (EO_TOU_007, no ellipsis in the German translation of this passage)

⁵² SFL term for a noun phrase

⁵³ For each corpus example, the ID of the respective corpus text is given.

The line drawn by Halliday and Hasan between verbal and clausal ellipsis is not very sharp. They state that it is possible to look at verbal ellipsis ‘from another angle, taking the clause as the point of departure’ (ibid.: 197), and to interpret it as a clausal ellipsis. Clausal ellipsis is a rather broad category. It is the omission of a part of a clause.⁵⁴ Clausal ellipsis often involves the omission of a clause to the exception of a single constituent, for instance in a question-answer sequence (4:5).

(4:5) a) *Wer sagt das? Meine Eltern.* (GO_FICTION_001)

b) *Who said that? My parents [].* (ETRANS_FICTION_001)

The three ellipsis subtypes nominal, verbal and clausal ellipsis occur both in written and spoken language in English and German with different register-specific frequencies. Halliday and Hasan’s argumentation on textual cohesion is based on the possibilities that exist in English, but we can generally apply these categories to German as well. Using such general categories precludes many of the flaws of more fine-grained existing typologies. It also helps to avoid the consequences of an overspecified model describing too many small samples in a quantitative analysis. Some ellipsis subtypes that are possible in English and / or German and that have been described extensively in the literature are

⁵⁴ It has sometimes been understood as the omission of an entire clause (Sutherland, 2016: 18), but the omission of a full clause would not be an ellipsis as it leaves no remnant structure of a clause.

relatively rare in actual language use, e.g. sluicing.⁵⁵ It seems reasonable to subsume them under more general categories in a corpus-based, quantitative analysis. Cohesive ellipses can be expected to be less frequent than various other types of cohesive devices. In fact, most cohesive links are probably lexical, and the resource used most often is lexical cohesion – in many cases repetition of the same item (Taboada, 2004: 172).

As the ellipsis types of nominal, verbal and clausal ellipsis are still an under-researched topic in studies on textual cohesion, we will address the question of what types of structures fall under these categories in English and German and how these are instantiations of cross-linguistic categories. There are some differences between the structure of noun phrases, verbal groups and whole clauses in English and German that may seem subtle with regard to some aspects and more striking with regard to others (inflectional morphology, word order, features of phrases and structural elements in phrases etc.). These cross-linguistic differences and similarities between English and German will be addressed for each category in the following chapters.

The next chapter will provide a detailed explanation of the cases that fall under nominal, verbal and clausal ellipsis in English and German. Therefore this section has limited its attention to a brief overview and description of these categories. Developing an ellipsis annotation scheme based on which manual annotation work will be carried out also has to deal with the issue of covering the wide variety of different omission

⁵⁵ e.g. *He invited someone, but I don't know who(m) [he invited] / Er lud jemanden ein, aber ich weiß nicht, wen [er einlud].*

possibilities in sentence structures of English and German texts from a broad spectrum of text types and communication scenarios on a written-spoken and formal-informal continuum. Therefore, the annotation scheme is rather detailed and gives numerous examples to cover a broad subset of the range of linguistic structures employed in naturally occurring text. The purpose of giving many examples is to provide a useful description that can be applied to real discourse data from the corpora and not only to idealised or prototypical isolated utterance pairs that we often find in the literature when ellipsis definitions are discussed. The next sections will explain the different types of ellipsis covered in the annotation scheme and discuss typical examples that are mainly taken from the GECCo corpus and that fall under the respective categories. Chapter 5 will also provide insights into the process of discussing and deciding on controversial cases in order to develop a consistent annotation standard. The ellipsis annotation scheme used for the annotation of ellipsis in the GECCo corpus and the ellipsis subcategories have also been sketched in Menzel (2014a; 2016b) and have been described and documented and in project-internal guidelines (Menzel, 2014b). The fine-grained ellipsis annotation guidelines that have been developed aim to ensure that all cases found in the corpus can be placed clearly in only one category in order to provide the basis for a meaningful quantitative analysis. The conceptualisation of ellipsis in this study combines several aspects of existing frameworks yet clarifies or adds others. It is based on theoretical research and considerations on ellipses in English and German as well as on the

occurrences of ellipses in an extensive and varied data base of our corpus
texts from a broad range of spoken and written registers.

5. Ellipses as cohesive devices

5.1 Nominal ellipses

5.1.1 The noun phrase in English and German

Before describing the types of nominal ellipsis covered by the annotation scheme, this section discusses the internal structure of the noun phrase in English and German. This is a necessary preliminary to any understanding of the possible structures of noun ellipsis remnants and the syntactic configurations that allow nominal ellipsis in these languages. The phenomenon of nominal ellipsis has been illustrated in Halliday and Hasan (1976: 147ff.). Nominal ellipsis can be defined as the omission of the head noun in a noun phrase with the purpose to avoid explicitly mentioning or repeating a noun. It is also sometimes called ‘head noun ellipsis’ in the literature (e.g. McShane, 2005: 128ff.), ‘NP-ellipsis’ (Corver and van Koppen, 2009) or simply ‘noun ellipsis’ (Sleeman, 1993). I view these terms as synonymous, although we avoid the term ‘NP-ellipsis’ for this phenomenon as this suggests that a whole phrase is omitted in a larger grammatically incomplete unit.

Omissions of other grammatically necessary elements from noun phrases such as articles do not fall under the category of nominal ellipsis. Grammatically acceptable article omissions are restricted to very specific syntactic environments such as instruction manuals or headlines of journalistic texts. They will be treated under the category of text-type-

specific fragments (cf. Chapter 6.5).

In English and German, prepositional phrases that consist of a preposition and a noun phrase can involve the omission of the noun phrase part while the underlying structure is anaphorically linked to another prepositional phrase in the text, either in the same sentence in coordinated structures involving contrast (5:1) or in clausal ellipsis in adjacency pairs such as question-answer sequences (5:2, cf. Chapter 5.3.2). Such cases where the preposition as the head of a prepositional phrase remains as remnant do not occur in the corpus of this investigation⁵⁶, and they would not fall under nominal ellipsis either.

(5:1) *You may have a choice of accommodation on campus or off [],
with meals or without [].*⁵⁷

(5:2) *Deine Suppe kommt gleich. Möchtest du sie mit Aioli oder lieber
ohne []? – Lieber ohne [].*⁵⁸

⁵⁶ (apart from one exophoric corpus example from a tourism text: ‘*Baden mit und ohne []*’ (GO_TOU_011) which has been translated literally as ‘*Bathing with and without []*’ (ETrans_TOU_011). German colloquial language allows such exophoric noun phrase ellipsis after prepositions (e.g. ‘*Punsch “mit und ohne []”*’).

⁵⁷ Example of noun phrase omissions in sentence-internal coordinated prepositional phrases, taken from the British National Corpus, EX5 W_non_ac_polit_law_edu.

⁵⁸ Example taken from the book *Lange Schatten im Oktober: Hallsteins zweiter Fall* (2014), Jürgen Heller, p.52,

<https://books.google.de/books?id=gp4HBgAAQBAJ&pg=PT52>

Similar structures can be found in English, but probably less frequently, e.g. ‘*Onion. Do you want your michigan with or without []?*’ – ‘*With [], please.*’ from the book *Letters in the Attic*, (2007), Bonnie Shimko, p.158,

<https://books.google.de/books?id=pVaXAAwAAQBAJ&pg=PT158> [14/03/2016]

According to Quirk et al. (1985: §2.27) all types of phrases are ‘groups’ of one or more words which can function as elements of clause structure. In English and German, phrase classes reflect major lexical word classes in noun phrases, verb phrases, adjective phrases, adverb phrases and prepositional phrase (the latter are noun phrases with a preposition added at the beginning). Each phrase has its focus on a head word of the relevant class that has certain potential extensions. A minority view holds that the determiner is the head of the noun phrase so that it should be considered a ‘determiner phrase’. Payne (1993) and Radford (1993) even suggested that the noun phrase has several heads, or is at least double-headed, as modified nominals incorporate multiple phrasal projections, with each modifier heading a separate projection. In the traditional view that we choose to follow here, the noun is the only syntactic and semantic head of the phrase.

We will use the terms ‘noun phrase’ / ‘nominal phrase’ and the SFL term ‘nominal group’ synonymously as is frequently done in the literature, although Halliday (1985: 159) made a distinction between the phrase as a contraction of a clause and the group as an expansion of a word. In the description of the internal structure of English and German noun phrases, we will leave aside those that consist of only a pronoun which replaces a noun (e.g. ‘*they*’, ‘*someone*’) as we are interested in those phrases that provide contexts for the potential omission of the head noun.

Abney (1987) described the English noun phrase in its sentential aspect. An overview of the structure of the English noun phrase from different

linguistic perspectives is given in de Mönnink (2000, Chapter 1); and Reich (2001) and Keizer (2007) also discuss the nature of the English noun phrase in great detail. There are several prominent publications on the noun phrase which treat this structure from a generative point of view (e.g. Alexiadou et al., 2007), but there is a diversity of approaches and terminology used to describe the internal structure of the noun phrase. Hill (1958: 176) claimed that nominal phrases have six positions in English with different ‘order classes’. Words belong to the same class if one word can replace the other without affecting the structure of the phrase. If different words belong to the same class, they can occur in either order. There is a normal, grammatically required unmarked order in English nominal premodifiers. Hill illustrated the noun modifier positions with the following phrase:

VI	V	IV	III	II	I	Noun
all	the	ten	fine	old	stone	houses

Figure 2: The structure of the noun phrase according to Hill (1958: 176, cited in Feist, 2008: 12)

Quirk et al. (1985) described ‘zones’ of use for the position of premodifiers in the noun phrase which may have one word, several, or none and which have certain semantic characteristics.

Determiners	Premodifiers				Head
	Zone I: precentral	Zone II: central	Zone III: postcentral	Zone IV: prehead	
<i>our</i>	<i>numerous</i>	<i>splendid</i>		<i>African tourist</i>	<i>attractions</i>
<i>all this</i>			<i>costly</i>	<i>social</i>	<i>security</i>
<i>a</i>	<i>certain</i>			<i>church</i>	<i>tower</i>
<i>these</i>			<i>crumbling[,] grey</i>	<i>Gothic church</i>	<i>towers</i>
<i>some</i>		<i>intricate</i>	<i>old[,] interlocking</i>	<i>Chinese</i>	<i>designs</i>
<i>all the</i>		<i>small</i>	<i>carved</i>	<i>Chinese jade</i>	<i>idols</i>
<i>both the</i>	<i>major</i>			<i>Danish political</i>	<i>parties</i>

Figure 3: Zones of premodification (approximate) according to Quirk et al. (1985: 1340, cited by Feist, 2011: 9)

Other authors have used more descriptive terms. Feist (2011: 56) used the terms ‘reinforcer’ for Quirk et al.’s zone I words, ‘epithet’ for zone II & III words, ‘descriptor’ for zone III and ‘classifier’ for zone IV (Figure 4). Noun-phrase complexity, the order of elements of the English noun phrase and word-order variation are comprehensively discussed by Feist (2008, 2011) and Berlage (2014), who analysed long groups of noun modifiers and addressed the question of why modifiers regularly occur in different positions. There are semantic and syntactic explanations of unmarked word order.

Determiner	Premodifiers			Classifier	Head
	Reinforcer	Epithet	Descriptor		
<i>your</i>	<i>actual</i>	<i>tinny</i>	<i>round</i>	<i>percussion</i>	<i>instrument</i>
<i>a</i>	<i>mere</i>	<i>useless</i>	<i>gibbering</i>	<i>stop-the-war-at-any-price</i>	<i>pacifist</i>

Figure 4: The structure of the English noun phrase according to Feist (2011: 56)

In Systemic Functional Linguistics, the positions and categories of ‘deictic’, ‘numerative’, ‘epithet’, ‘classifier’ and ‘qualifier’ have been suggested by Halliday and Hasan (1976: 147) for pre- and postmodifiers of nouns, illustrated by the phrase: ‘*those* (deictic) *two* (numerative) *fast* (epithet) *electric* (classifier) *trains* (head noun) *with pantographs* (qualifier)’. A more detailed distinction between the elements and positions from an SFL perspective is made in Matthiessen et al. (2010), cf.

Figure 5.

<i>a picture of</i>	<i>these</i>	<i>famous</i>	<i>first</i>	<i>two</i>	<i>marvellous</i>	<i>brick</i>	<i>houses</i>	<i>with gardens</i>
Facet	Deictic	Post-Deictic	Ordinative	Numerative	Epithet	Classifier	Thing	Qualifier
nom.gp	determiner	adjective	numeral: ordinal	numeral: cardinal	adjective	noun	noun	prep. phrase

Figure 5: The structure of the nominal group according to Matthiessen et al. (2010: 150)

The differences between the English and German noun phrase which lead to slightly different possibilities for nominal ellipsis are mainly morphological differences in the inflectional case system. Teich

(2003: 125) suggested that apart from differences with regard to the inflection paradigm, the major “systems” of the nominal group (e.g. the system of specificity or the system of nonpronominal types including numeration, ordination, epithesis, classification and qualification) are shared between English and German. The internal structure and the syntactic features of the German noun phrase have been described, for example, in Haider (1988, 1992), Bhatt (1990) and Zifonun et al. (1997: 1926ff.). As a general proposition it can be stated that – despite the well-known differences with regard to inflectional morphology of nouns and noun phrase elements such as determiners, numeral or adjectives – the structure and phrasal architecture of the noun phrase in German has many similarities to its English counterpart. In König and Gast’s textbook on English-German contrasts (2012: 208f.), we find a brief overview on the English and German noun phrase and the statement that the order of constituents within the German and the English noun phrase are ‘by and large identical’. Nevertheless there are some subtle differences in the structure of English and German noun phrases that are worth pointing out as they may have an influence on contexts where nominal ellipsis is possible and lead to slightly different structures that can be left as ellipsis remnants.

Noun phrases in English and German often have a pre-modifier, usually a determiner (article, demonstrative or possessive). A noun phrase can usually have only one determiner. There is a small number of pre-determiners in English, most of them concerned with quantification which

tend to occur before determiners (e.g. ‘*all [of] my books*’, cf. Jeffries, 2006: 105). We find multiple determiners in the form of a combination of a demonstrative or article and a possessive in some languages (e.g. in Romance languages such as Italian ‘*questa mia bella città*’ [**this my beautiful city*] or ‘*un mio amico*’ [**a my friend / a friend of mine*]) where the adjectival character of possessives is stronger than in English. Such a combination is generally not permitted in modern English.

In English and German as well as in other Germanic languages, combinations of demonstratives and possessives were sometimes used in older stages of the respective languages where possessives sometimes tended to function more like an adjective and less like a determiner. For instance, there are examples in the Bible that sound archaic in modern English and German, e.g. ‘*this my son*’ (Luke 15:24, King James Version)⁵⁹, ‘*dieser mein Sohn*’ (Lutherbibel, 1912). There are similar structures in the Norwegian, Danish, Swedish translation of this passage (‘*denne min sønn/søn/son*’) which can be explained by the use of both a demonstrative and a possessive in the Latin and Greek bible texts from which the translations were made: ‘*hic filius meus*’ (Vulgate), ‘*οὗτος ὁ υἱός μου*’⁶⁰ (this the son of-me [i.e. ‘*mine, my*’])⁶¹. In modern German, we

⁵⁹ Bible passages cited from <https://www.biblegateway.com> or <http://bibeltext.com> if not indicated otherwise [06/01/2016]

⁶⁰ <http://www.scripture4all.org/OnlineInterlinear/NTpdf/luk15.pdf> [06/01/2016]

⁶¹ The Greek text has a definitive article in this noun phrase while Latin generally did not have articles. Interestingly, none of the Romance languages can have a structure such as the Latin demonstrative – noun – possessive in ‘*hic filius meus*’. Portuguese, Spanish and Italian have the following structure: demonstrative – possessive – noun: ‘*este meu filho*’ (Almeida Revista e Corrigida, 2009) / ‘*este mi hijo*’ (Reina Valera, 1909) / ‘*questo mio figlio*’ (Nuova Riveduta, 1994). In French, the demonstrative element comes after the noun in the form of a relative clause (‘*mon fils que voici*’, Nouvelle Edition de Genève, 1979) and in Romanian in which the position of the elements resembles Latin most, we

sometimes find multiple determiners in front of a noun. They sound formal and can be used in emphatic and poetic language as in *'diese meine Absicht'* (GO_POPSCI_004) or in the phrase that Helmut Kohl frequently used: *'dieses unser Land'*⁶². Additionally, German can have combinations of demonstrative articles and possessives which do not sound archaic, but slightly formal (*'Wessen Schal ist das? Ist es der meine?'⁶³* instead of the unmarked form *'Ist es meiner?'*). Even more formal possessive forms exist combining determiner and possessive as in *'Unser Bauplatz ist dicht bei dem eurigen.'*⁶⁴

These combinations of articles and possessives look like nominal ellipsis at first sight, but they can also be analysed differently. A slightly similar discussion can be observed in the literature with regard to Dutch. In contrast to German, possessives after determiners are the usual, unmarked forms in standard Dutch if the head noun has not been overtly realised (e.g. *'mijn vader'* → *'de mijne'*).⁶⁵ Corver and van Koppen (2007) analyse these constructions as remnants of noun ellipsis. It would also be possible to assume that the possessive *'mijne'* is not part of an ellipsis construction but the head of the noun phrase as a nominalised possessive

find the 'AL-construction' with AL as an additional word as possessive marker (*'acest fiu al meu'* [Cornilescu, 1924]).

⁶² cf. for instance the speech "Über die Zukunft der deutschen Demokratie" 15/01/1970 <http://www.ueberseeclub.de/resources/Server/pdf-Dateien/1970-1979/vortrag-1970-01-15Dr.%20Helmut%20Kohl.pdf> [last checked 06/01/2016]

⁶³ Example taken from the Duden: <http://www.duden.de/rechtschreibung/meine> [last checked 19/02/2016]

⁶⁴ Example taken from the Duden <http://www.duden.de/rechtschreibung/eurige> [last checked 19/02/2016]

⁶⁵ Some varieties such as Winterswijk Dutch or Hindeloopen Dutch tend to use possessives without determiners, periphrastic possessives or constructions that resemble the German informal or dialectal possessive dative construction (nounDAT + possessive 3rd person sg + nounNOM, e.g. *'dem Vater seine Häuser'* → *'dem Vater seine []'*).

(Broekhuis and den Dikken, 2012: 840).

In English, the use of determiner-possessives sequences is more restricted. Such patterns do occur occasionally in larger corpora in artistic or fictional language when a noun follows, for instance, in this extract from a fantasy novel which is cited in the British National Corpus:

(5:3) *This my spell on you is laid.* (BNC, F99 W_fict_prose)

As determiner-possessives sequences are archaic and very formal or ungrammatical nowadays, we do not find them in our corpus in contexts that would licence nominal ellipsis and it would be beside the point to discuss them in more detail. These examples were cited primarily to illustrate a subtle difference between the structure of the English and German noun phrase.

There are a few other differences between the structure of the English and German noun phrase with regard to the elements that can precede or follow a noun within the noun phrase and with regard to the frequency for options that generally exist in both languages.

In some cases, although mainly in archaic and poetic uses or in phrases borrowed from Romance languages, English allows post-nominal non-modified adjectives (e.g. *'faith in things unseen'*, *'heir apparent'*, *'God Almighty'*, *'pound sterling'*, *'Alcoholics Anonymous'*). Some English adjectives can be found pre- or postnominally (e.g. *'adjacent'*, *'present'*, *'responsible'*). The meaning of certain adjectives depends on their

position. Only in very few cases does an adjective without further extension occur after the noun (5:4-5).

(5:4) *'Look for the Shetland ponies in the field adjacent!'*
(EO_TOU_002)

(5:5) *Recruiting the best people available, UPS merged a number of cultures and procedures into a seamless operation called UPS Airlines.* (EO_WEB_009)

Such postnominal adjectives without complements can be interpreted as reduced relative clauses, similarly to postnominal adjectives with complements (5:6-7).

(5:6) *It illustrates a principle fundamental to the EU's approach to its neighbourhood.* (EO_ESSAY_004)

(5:7) *I began to adopt a frame of mind appropriate for the journey before me.* (EO_FICTION_006)

Other examples of noun phrases where an adjective comes after the noun in the corpus data are Latin medical terms or names of species. It is not considered necessary to describe such constructions here in great detail

as they are a marginal phenomenon in our corpus data.⁶⁶ English can also have adjectives after indefinite pronouns (e.g. *'something special'* = some special thing / something that is special or *'someone strong'* = some strong person / someone who is strong. They correspond to nominalizations in German.⁶⁷

In general, we assume that English tends to use post-modifiers and complements such as prepositional phrases or clauses that come after the head noun that often correspond to a premodifying structure in German. Nevertheless, English does permit some extension of premodifiers that are unusual in German, e.g. long hyphenated phrases, sometimes used in ad-hoc compounds, which can be creatively used as embedded phrases and clauses in the position of a noun modifier in English (e.g. *'a once-in-a-generation opportunity'*, EO_ESSAY_029). English compounds often consist of several orthographic words and can pose problems of delimitation between a nominal phrase and a compound. Adjective + noun constructions, for instance, can lead to an omission of an orthographic word as a part of the compound (5:8) which is slightly different from nominal ellipsis after an attributive adjective in a noun phrase.

⁶⁶ It is also difficult to find such examples with corpus queries. Due to the non-standard syntax, adjectives after nouns, especially rare words derived from Greek or Latin, are sometimes wrongly tagged as a noun in the GECCo corpus, as for instance in *'Caenorhabditis elegans'* (EO_POPSCI_003)

⁶⁷ <http://www.duden.de/sprachwissen/rechtschreibregeln/Gro%C3%9F-%20und%20Kleinschreibung#K72> [last checked 19/02/2016]

(5:8) *The signs of both roe deer and red [] are found throughout the wood.*⁶⁸

In German, there are many possibilities for placing present or past participles or extended participle constructions in front of a head noun to function as adjectival modifiers which is possible in fewer cases in English, and there mainly for past participles and only for a few present participles (5:9-5:13). English will often have to put equivalent constructions after the noun (5:14-5:15). Nevertheless, English does have the possibility of creatively constructing adjectives from participles, particularly to form compound adjectives or phrasal compounds (e.g. ‘*New York-based*’, ‘*slow-moving*’, ‘*home-cooked*’, ‘*urgently-needed*’), that can result in long noun modifiers (e.g. ‘*a London-born second-generation Irishman*’⁶⁹, ‘*the well-resourced, self-contained services in the developed sector*’⁷⁰, ‘*a never-seen-before constitutional crisis*’⁷¹, ‘*a long-awaited and never-to-be-forgotten hug*’⁷²), but compound adjectives and phrasal compounds are relatively rare in most registers included in our corpus, apart from fictional and popular scientific texts, and we assume that they

⁶⁸ Example taken from: <http://www.woodlands.co.uk/pdf/hawhill-wood.pdf> [last checked 19/02/16]

⁶⁹ Example from the book *Sports Events, Society and Culture*, Katherine Dashper et al. (eds.) (2014), p.139, <https://books.google.de/books?id=2xAWBAAAQBAJ&pg=PA139> [19/02/2016]

⁷⁰ Example from *Comparative Studies in Special Education*, Kas Mazurek (1994), p.17, https://books.google.de/books?id=zOl_hpr0MiEC&pg=PA17 [last checked 19/02/2016]

⁷¹ From *Southeast Asia in a New Era: Ten Countries, One Region in ASEAN*, Rodolfo C. Severino et al. (eds.) (2010), p.208, <https://books.google.de/books?id=NcV6BwAAQBAJ&pg=PA208> [last checked 19/02/2016]

⁷² Example from the novel *The Gift of a Legacy*, Jim Stovall (2013), p.190, <https://books.google.de/books?id=SmJ2Mok64p0C&pg=PA190> [last checked 19/02/2016]

generally cannot function as noun ellipsis remnants.

(5:9) *the hoped-for economic growth / das erhoffte
Wirtschaftswachstum* (EO/GTRANS_ESSAY_026)

(5:10) *interesting locally made cheeses / interessante, vor Ort
produzierte Käsesorten* (EO/GTRANS_TOU_002)

(5:11) *the newly appointed judges / die neu ernannten⁷³ Richter*
(EO/GTRANS_SPEECH_004)

(5:12) *heather covered heights / heidebestandene⁷⁴ Höhen*
(EO/GTRANS_TOU_006)

(5:13) *the soil-dwelling worm / der im Boden lebende Fadenwurm*
(EO/GTRANS_POPSCI_003)

(5:14) *the firms working on these projects / die an diesen Projekten
arbeitenden Unternehmen* (EO/GTRANS_SPEECH_004)

(5:15) *the plus symbol adjacent to the group / das nebenstehende
Plussymbol* (EO/GTRANS_INSTR_004)

⁷³ Also sometimes spelt as one word in German ('*neuernannt*')

⁷⁴ The use of this compound adjective is slightly marginal in German. The components of such compounds can often be used as separate words + preposition to modify the noun ('*mit Heide bestandene Höhen*').

As German participles can be inflected as adjectives are and agree in gender, number and case with the noun in German, the head noun can be omitted if its meaning is clear from the context (5:16, 5:17a, 5:18a). If an equivalent participle exists that can be used as a modifier in English, usually the substitute ‘one’ (5:17b) or repetition, a synonym another head noun (5:18b) has to be used to morphologically mark number and to disambiguate the function of the participle. Nominal ellipsis after a participle adjective occurs in very few examples in our corpus data, e.g. (5:19), and it can display syntactic interference from German in English translations (5:20).

(5:16) *Der Schaum hält sich genauso lange wie der per Aufschäumgerät erzeugte [] und ist auch genauso feinblasig. (GO_FORUM_010)*

(5:17) a) *Dieser Großblick auf das Weltsystem wird wiederum durch das relativiert, was der James Rosenau die „zwei Welten der Weltpolitik“ nennt, nämlich durch die Vorstellung, daß es nicht eine globale Gesellschaft gibt, sondern mindestens zwei miteinander konkurrierende []. (GO_POPSCI_002)⁷⁵*

b) *But this world-system view has in turn been nuanced by reference to what James Rosenau calls 'the two worlds of world politics': that is, the idea that there is not a single global society but at least two competing ones. (ETRANS_POPSCI_002)*

⁷⁵ The letters a) and b) are used in sample sentences from the corpus that are sentence pairs where one sentence is the translation of the other.

(5:18) a) *Dem Selbstbewußtsein vermag er außer der autoritären Seite eine versöhnende [] nicht mehr abzugewinnen.*

(GO_POPSCI_001)

b) *He can no longer glean from self-consciousness any reconciling dimension in addition to its authoritarian aspect.*

(GO_POPSCI_001)

(5:19) *The Forestry Commision, the government body responsible for state forestry have also developed a very extensive choice of forest paths from the undemanding to the fairly testing [].*

(EO_TOU_002)

(5:20) * *The term “spa” is also the attached [] to the seaside wellness resorts.⁷⁶* (ETRANS_TOU_010)

We occasionally find examples of nominal ellipsis in our German corpus data where certain remnant constituents occur in entirely different phrase structures in English, for instance when an attributive adjective corresponds to a predicate adjective in the translation (5:21).

⁷⁶ Graded (un)acceptability is indicated by a question mark in front of a sentence and ungrammatical or non-existing sentences are marked with an asterisk (*).

- (5:21) a) *Zunächst sollen solche Ansätze zur Sprache gebracht werden, die jeweils eine spezielle Dimension und Logik der Globalisierung zur zentralen [] erklären.* (GO_POPSCI_002)
- b) *First, we should consider approaches which hold one special dimension or logic of globalization to be central.*
- (ETRANS_POPSCI_002)

The order of adjectives in noun phrases can sometimes differ between originals and translations due to semantic or stylistic preference. Not only is there a general tendency to place long or ‘heavy’ noun phrases at the end of a clause, but also to place longer noun phrase elements towards the end of a noun phrase if there are no semantic constraints on their order (5:22).

- (5:22) *dein kleines hässliches japanisches Auto / your ugly little Japanese car* (GO/ETRANS_FICTION_005)

Doherty pointed out some differences with regard to nominal groups in English and German that result in potential translation difficulties. In her opinion, prenominal modification is generally restricted in English while postnominal modification is restricted in German, which has information-structural effects (Doherty, 2006: 74). Due to the richer inflection paradigms of possible constituents of noun phrase remnants in German, we can expect nominal ellipsis to occur in more contexts in German where an equivalent elliptical structure is not possible in English. Additionally, we

assume that German with its highly nominal style (Kunz 2010: 172) is characterised by a higher number of noun phrases and generally by longer nominal groups with more possibilities to obtain elliptical noun phrases.

5.1.2 Cohesive and non-cohesive cases of nominal ellipses

Several analyses of nominal ellipsis have been proposed (e.g. Lobeck, 1995 for English; Giannakidou and Stavrou, 1999, Ntelitheos, 2004, Alexiadou and Gengel, 2012 for Greek; Kester and Sleeman, 2002, Eguren, 2009 for Spanish; Sleeman, 1996, Valois et al., 2009 for French). Llombart-Huesca (2002) compared anaphoric *'one'* and nominal ellipsis in English and Günther (2011, 2013) described the general structure of elliptical noun phrases in English and the different types of licensors of empty nouns in English with a brief comparison to the possibilities in German. She particularly focused on nominal ellipsis after adjectival modifiers in English corpus data. Often nominal ellipsis involves the deletion of nouns after modifiers such as descriptive adjectives, numerals or adjectives expressing indefinite quantities. In contrast to other ellipsis types, we will presumably find nominal ellipses in all registers of spoken language as well as in various written text types such as narrative, technical or business writing.

If a nominal ellipsis has a textual antecedent, it is recoverable or inferable from the co-text, i.e. it signals to hearers or readers that they should carry over the wording from another noun phrase in the surrounding text. Endophoric nominal ellipsis can occur as a clause-internal phenomenon or even within a phrase. In those cases, it is not primarily used as a cohesive device. If the antecedent of a nominal ellipsis is in the same clause or phrase, which is quite often the case, the effect of textual cohesion is less significant. Clause-internal anaphoric ellipses occur, for instance, after linking verbs in incomplete predicate noun

phrases (5:23a), in appositions (5:24) or in the second conjunct of a coordinated noun phrase (5:25-27).

(5:23) *a) Our economy is one of the most productive [].*

(GO_SPEECH_014)

b) Unsere Wirtschaft gehört zu den produktivsten [] überhaupt.

(ETRANS_SPEECH_014)

(5:24) *Wir haben ein großartiges Unternehmen, eines der weltweit erfolgreichsten [].*

(5:25) *Can I ask a question or two []? (EO_FORUM_007)*

(5:26) *Equipping people for the jobs of today's world, and tomorrow's [], is essential in the global economy of the 21st century. (EO_ESSAY_004)*

(5:27) *Die heutige Abendvorstellung und die morgige [] sind komplett ausverkauft.*

Clause-internal nominal ellipses are annotated separately to clearly distinguish them from cohesive ellipses and to compare their frequency with that of ellipsis-antecedent relations that go beyond clause or sentence boundaries. In some cases, whole clauses may have been reduced

elliptically, which has to be considered in the annotation of cross-clausal ellipses. In the annotation, clause-internal ellipses are linked to their nearest clause-internal antecedent and not to additional clause-external identical nouns as antecedents (5:28). This would only be necessary if such nouns occur in the nearer co-text (about one or two sentences) and play a major role for the interpretation of the ellipsis. Usually, clause-internal ellipses can be resolved without information from previous sentences.

(5:28) a) *She poured a scalding kettle of water* (not annotated as antecedent) *into the sink*. 'Why didn't you go to bed?' I asked her. 'No point if I had to get up with you three hours later.' She shot a jet of cold water ('water' annotated as clause-internal antecedent) *into the hot [water]*. (EO_FICTION_008)

b) *Sie kippte einen Kessel mit kochend heißem Wasser* (not annotated as antecedent) *in den Spülstein*. „Warum bist du denn nicht ins Bett gegangen?“ fragte ich. „Weil es sich nicht gelohnt hat, wo ich drei Stunden später wieder mit dir aufstehen muss.“ *Sie ließ einen Strahl kaltes Wasser* ('Wasser' annotated as clause-internal antecedent) *in das heiße [Wasser] laufen*. (GTRANS_FICTION_008)

It may be argued that nominal ellipsis in coordinated phrases in general is a type of non-standard word order or extraposition – and no omission – where, for example, one part of a phrase is moved towards the end of a

phrase in order to achieve metrical prominence and focal emphasis (e.g. *'die großen Fische und die kleinen'*, *'this year and next'*).

As English has different word formation rules for compounds, we find hyphenated two-word compounds used as adjectives (e.g. *'below-market rates of return'*, 5:29a) that can become rather long noun modifiers in the German translation. Due to the freer word order in German, translators can move long constituents or parts of such a long constituent towards which results in structures that may either be seen as an extraposition or an ellipsis.

(5:29) a) *And if the EU does as it has in the past, and provides financing to Airbus at below-market rates of return, we could be facing a very large and highly contentious fight in the WTO.*
(EO_SPEECH_009)

b) *Und wenn die EU sich wie in der Vergangenheit verhält und dem Airbus Finanzierung zu Zinssätzen unter den auf dem Markt gültigen [] bietet, könnte uns ein großer und sehr kontroverser Kampf in der WTO bevorstehen.*⁷⁷ (GTRANS_SPEECH_009)

In my corpus analysis, these structures fall under non-cohesive nominal ellipsis. In any case, these are clause-internal phenomena which are less important for the analysis of textual cohesion.

⁷⁷ 'zu Zinssätzen unter den auf dem Markt gültigen' = 'zu unter den auf dem Markt gültigen Zinssätzen' ?

The examples given above describe endophoric relations, but a nominal ellipsis may also refer exophorically to extra-linguistic elements that can be recovered from the situational context. These ellipses do not have textual antecedents and are not used endophorically. If no syntactically matching antecedent is available, other factors such as knowledge about the situation, the text type and world knowledge come into play for the interpretation of the ellipsis:

(5:30) *I'll take four [].* (+ pointing gesture)

(5:31) *Soll ich das rote [Auto] oder das blaues [Auto] kaufen?* (uttered by someone pointing at two cars)

Some nominal ellipses are relatively independent of the co-text and context as they have become conventionalised or lexicalised ('lexikalische Ellipse', Ágel, 1991). They have a standard interpretation and need no supporting head noun (e.g. *go to a friend's [house], Macy's [department store], a barber's [shop], the Baltic [Sea], the Mediterranean [Sea]*).⁷⁸

Exophoric and lexicalised nominal ellipses, which we observe only occasionally in our data, and clause-internal nominal ellipsis will be subsumed under the category of non-cohesive ellipsis in the analysis. Nominal ellipses are annotated as cohesive ellipses if a relation between a

⁷⁸ English lexicalised ellipses which are the result of shortened adjective noun combinations – and English adjective noun combinations in general – often correspond to noun + noun compounds in German, which explains why we find fewer of them in German, e.g. *Mediterranean [Sea] / Mittelmeer*.

nominal ellipsis and its antecedent occurs across sentences. The distance in words between the ellipsis site and the antecedent in a cross-sentential relation is usually not very large. If the antecedent and the ellipsis are not embedded in perfectly parallel structures, for instance, when they do not both occur in subject position at the beginning of different sentences, they tend to be separated only by a few words or by a punctuation mark. (5:32-33), for instance, have an elliptical noun phrase at the beginning of the sentence while the antecedent noun phrase is directly at the end of the previous sentence.

(5:32) *Almost every district has its own open-air market. The largest and most impressive [] is Naschmarkt near the Secession building.*
(ETTRANS_TOU_022)

(5:33) *Das Muster an Genaktivitäten und erzeugten Proteinvarianten, auf dem diese erstaunliche Kreation basiert, hängt stark von zwei Mechanismen ab. Der erste [] besteht in einer Modifikation des Chromatins.* (GTRANS_POPSCI_004)

Nominal ellipses are also annotated as cohesive ellipses if a relation between a nominal ellipsis and its antecedent occurs across coordinated or subordinated clauses (5:34). This type of ellipsis contributes to the cohesiveness of a text even if it occurs across clauses within the same sentence. Cross-clausal nominal ellipses will be included in the analysis of

cohesive cases and will not be regarded as a mere result of coordination. They are not strictly bound to the coordinated or subordinated clausal structure and the ellipsis would usually also be possible if it was anaphorically related to an antecedent in a previous sentence or if an additional sentence was put between the two clauses.

- (5:34) a) *On one side of the road, thickets and small trees rose steeply, while on the other [] I could now glimpse through the foliage the distant countryside.* (EO_FICTION_006)
- b) *Auf der einen Straßenseite zogen sich Dickicht und kleine Bäume steil aufwärts, während ich auf der anderen [] durch das Laub jetzt die ferne Landschaft herüberschimmern sah.* (GTRANS_FICTION_006)

Nevertheless, even when a nominal ellipsis is used to link clauses, pairs of sentences or longer passages of texts as a cohesive device, it would not be accurate to say that such an ellipsis is entirely independent of its immediate syntactic environment. Clauses that are linked by an ellipsis-antecedent relation often have similar structures that lead to a certain effect of structural priming (Xiang et al., 2014). Example (5:34), for instance, involves certain parallelisms between the prepositional phrases ‘*on the one side of the road*’ – ‘*on the other []*’ which both occur at the beginning of similarly structured clauses. The German translation ‘*auf der einen Straßenseite*’ – ‘*auf der anderen []*’ is embedded in a different word

order pattern where the subject either precedes or follows the prepositional phrase and where the finite verb comes directly after the prepositional phrase in the main clause on the one hand and at the end of the sentence in the subordinate clause on the other hand. In both languages, there is an effect of structural priming between the complete and the elliptical prepositional phrase. Identical case marking of the antecedent phrase and the ellipsis remnant indicates parallelism in German while English here shows the tendency to compensate for the lack of case morphology with stricter parallel word order in ellipsis-antecedent constructions. Additionally, as the second constituent expresses some semantic contrast to the first constituent ('one' / 'einen' vs. 'other' / 'anderen'), both elements will usually be marked by contrasting prosodic signals in spoken language. A contrastive, complimentary or sequential relationship is usually indicated by noun modifiers in the antecedent and the ellipsis remnant, e.g. 'one/another', 'first/second'. The relationship between contrasting, complementary or sequential constituents of which one involves a nominal ellipsis can be enforced by other explicit signal words such as conjunctions ('both/and', 'not only/but', 'either/or' etc.). (5:35) is another example with a combination of parallelism and contrast in two clauses of which one has an elliptical noun phrase.

(5:35) *U.S. citizens don't hold foreign currencies here, they just hold domestic [].* (EO_ACADEMIC_003)

Parallel structures in which grammatical patterns are repeated in series of sentences have a strong cohesive effect and have been suggested as a specific means of grammatical cohesion (Schubert, 2008: 40ff., de Beaugrande and Dressler, 1981: 58). Parallelisms can involve, for instance, a series of similarly structured elements having the same length (isocolon). The repetition of morpho-syntactic surface structures in parallel patterns aligns related concepts so that lexical material embedded in parallel structures does not always have to be explicitly reiterated. Although there seems to be no strict adjacency requirement for nominal ellipsis, it usually occurs in close adjacency to its antecedent noun phrase.

In contrast to co-reference where a pronoun and its antecedent have to show number agreement, the grammatical number of the antecedent (and grammatical case, particularly in German) of an ellipsis may be different from that of the noun which could be inserted into the ellipsis site, but the antecedent and the omitted element have to be the same lexeme. If the grammatical number of the antecedent is different from that of the omitted noun, this has sometimes been discussed as a singular-plural mismatch (cf. Chapter 3.3) and some people's intuitive judgments may suggest that there is a cline in acceptability of such omissions. Hoffmann (1999) and Nunes and Zocca (2009) confirm that identity with regard to number is not a necessary condition in German and English, which is probably due to the fact that the omitted structure is not present on the surface of the text. Ellipses with 'flawed antecedents' from our corpus data will briefly be addressed in Chapter 9.2. In both languages, we find some nominal

ellipses with number mismatches referring to plural antecedents in most cases. We can generally expect number mismatches when a lexeme is omitted, repeated or substituted more than once so that we for example find cases where a nominal ellipsis refers to an antecedent lexeme that has been mentioned in singular and plural in the text or where several nominal ellipses are related to the same antecedent (5:36).

(5:36) *Thank you for posting such an easy and delicious recipe for basic yellow cake. I have tried so many [], and yours is the easiest [] I've come across.* (EO_FORUM_010)

Non-identity of number typically occurs when the ellipsis site and the antecedent noun phrase are related by a contrast between one specific entity and a larger group of entities, e.g. with a cardinal or ordinal number (5:37-38) in the ellipsis remnant.

(5:37) *Zimbabwe has – uh – its own, has a single field trial. They've got about fifty [], the Netherlands.* (EO_ACADEMIC_009)

(5:38) *Strandkörbe sind übrigens eine deutsche Erfindung, der erste [] wurde bereits 1882 gebaut.* (GO_TOU_011)

If ordinal numbers occur in both the elliptical and the antecedent noun phrase, they usually do not contrast a singular and a plural noun (5:39).

(5:39) ... our store of phrases that had been funny on first utterance and seemed scarcely less funny on the hundredth [...].⁷⁹

We can also expect cases of non-identity of number to occur more frequently in spoken language and in contexts where the ellipsis-antecedent relation stretches at least across clause boundaries.

(5:40) *Ehm, wenn Sie an die Informatik denken, in der Informatik gibt es die Programmiersprachen. Da kann man sich jetzt drüber streiten, welche die beste [] ist und welches die beste zu Grunde liegende Methodik ist. (GO_ACADEMIC_001)*

Certain specific structures subsumed by Halliday and Hasan (1976) under nominal ellipsis will not be included in my analysis. I will not annotate phrases as being incomplete when it is not possible to add any material without changing the form of other sentence elements or obtaining a sentence with a different meaning.

There are minor differences between rather similar structures in German and English, for example with regard to the pronouns / determiners ‘none’ / ‘keine(*r,s*)’ and possessives with inflectional suffixes or in non-reduced forms, e.g. ‘mine’ / ‘meine(*r,s*)’ and ‘yours’ / ‘deine(*r,s*)’. English and German determiners with inflectional paradigms that allow the insertion of a noun after some agreement suffixes are only

⁷⁹ Example from the novel *Purity* (2015) by Jonathan Franzen, p.392, <https://books.google.de/books?id=cT94BwAAQBAJ&pg=P392> [last checked 05/03/2016]

annotated when a noun insertion is possible and they have been marked as problematic in the annotation. Kunz and Steiner (2013) tend to analyse cases such as ‘*eine(r,s)*’, ‘*keine(r,s)*’, ‘*solche(r,s)*’, ‘*welche(r,s)*’ as nominal substitutes in German but state that some forms show structural ambiguity between substitution and ellipsis. Additionally, in contrast to Halliday and Hasan’s analysis, English demonstratives and their German equivalents (5:41) will not be regarded as grammatically defective or incomplete noun phrases, but as replacements of nouns even if they can be followed by an additional noun (Halliday and Hasan, 1976: 157)⁸⁰.

(5:41) *First of all we’ll take the starting point element and store that in ‘temp’.* (EO_ACADEMIC_006)

(5:42) *Ihr Haar war im Unterschied zu dem der anderen unter einem Schleier verborgen.* (GTRANS_FICTION_002)

Certain specific cases of nominal ellipsis occur in highly formulaic structures (e.g. *on the one hand – on the other [] / to name only a few [] / um nur einige [] zu nennen*). It would be possible to exclude such cases of nominal ellipsis completely from the analysis as they occur relatively frequently in the corpus and may have a strong influence on the interpretation of the results. I have chosen to include them in my analysis

⁸⁰ For some grammarians the potential multiple word-class interpretations of other elements that I chose to analyse as (albeit problematic) types of ellipsis remnants, such as quantifiers or possessives, would probably be an argument to see such elements also as proforms and replacements of the noun and not as noun ellipsis remnants.

as, in principle, their syntactic structure does not distinguish them from other nominal ellipses.

5.1.3 Nominal ellipses after numerals

This section describes ellipsis after numerals in English and German as cases that are widely accepted in the literature as being incomplete noun phrases.

Ellipsis after numerals in English⁸¹

- (5:43) *There are approximately 130 free trade agreements in force globally, the United States is a party to just two [free trade agreements]: one [free trade agreements] is with Canada and Mexico (NAFTA) and the other [free trade agreements] [is] with Israel. (EO_ESSAY_015)*
- (5:44) *This copy is defective but the other two [copies] are fine.*
- (5:45) *We have two keys but we need three [keys].*

⁸¹ Not all of the examples of ellipses that are given in this section on nominal ellipses after numerals are cohesive cases. I also include some clause-internal cases for a general illustration of the different types of ellipses from Sections 5.1.3-5.1.6. These types can be used as cohesive devices, but they are not always used in this function. This depends on the surrounding syntactical structures (cf. Section 5.1.2 on the distinction between cohesive and non-cohesive cases.)

Ellipsis after numerals in German

(5:46) *Obwohl es weltweit etwa 130 Freihandelsabkommen gibt, sind die Vereinigten Staaten nur zwei [] beigetreten: Einem [] mit Kanada und Mexiko (NAFTA) und einem [] mit Israel.*⁸²

(GTRANS_ESSAY_015)

(5:47) *Im Erbfall hätte unsere Tochter viel mehr bekommen als seine drei [Töchter].*⁸³

(5:48) *Vor den Spielen 2008 lagen zwischen EM und Olympia viereinhalb Monate, diesmal sind es nur zwei [Monate].*

(5:49) *Das Hotel hat zwar nur drei Sterne, hätte aber vier [Sterne] verdient.*

There is a difference between nominal substitution – where ‘one’ clearly substitutes the head noun – and nominal ellipsis after ‘one’ either as a numeral (5:50) or a full form of the indefinite article. In (5:51) ‘one’

⁸² According to the normative German Duden Grammar, the status of the cardinal number ‘eins’ (one) with regard to its word class is not clear. It combines features of adjectives, determiners and pronouns (cf. *Duden Grammar*, 2009, Chapter 3.7.2.2 on ‘Flexion der Kardinalzahladjektive’: 383). Cardinal numbers used as adjectives apart from ‘eins’ only show ‘rudimentary’ inflection. In several cases, the inflectional ending is optional and may sound more formal (e.g. Dativ: *mit zwei/zweien, mit drei/dreien*). After some inflectional endings of numbers the direct insertion of noun is not possible as in: ‘*Seraphim standen über ihm; ein jeglicher hatte sechs Flügel: mit zweien [/] deckten sie ihr Antlitz, mit zweien [/] deckten sie ihre Füße, und mit zweien [/] flogen sie*’. (*Lutherbibel*, 1912, Isaiah 6:2)

⁸³ Some examples in this chapter have been cited from various online resources, such as online discussion forums, and not from our corpus data.

can be regarded either as a numeral (the complete structure being ‘...*there may well be one common strategy*’) or the form taken by the indefinite article when it is functioning as the remnant of an elliptical noun phrase (‘...*there may well be a common strategy*’).

(5:50) *When asked for “one fish,” the children would hand the experimenter exactly one [], but failed to show any consistent interpretation of any larger numbers.*

(5:51) *Indeed it has been suggested this morning that there is a need for a common strategy among NATO nations. One would observe that there may well be one [].* (EO_SPEECH_013)

The borderline between substitution and other cohesive devices is not entirely clear from the literature. Halliday and Hasan distinguish the replacement of lexical items by grammatical substitutes (e.g. ‘*one*’ for nouns, ‘*do*’ for verbs or ‘*so*’ for clauses) from replacements by other pro-forms and consider pronouns as reference items where links between these items and their antecedents are established via co-reference. Quirk et al. (1985: 75ff., 854ff.), however, counted more forms as substitutes and explained that pro-forms are either related to their antecedent by co-reference or substitution, two concepts that are different in theory but overlap in practice (ibid: 803).

In cases where a numeral or another modifier such as an adjective or a

quantifier is accompanied by ‘*of them / of which / of whom / among them / davon / unter diesen*’ or similar structures it will not be analysed as an ellipsis here as it can be argued that those additional words in a partitive construction replace or substitute the omitted noun in the same noun phrase and it would be very unusual to repeat the noun in such cases (5:52-5:54). Partitives after quantifiers (as well as some other constructions with quantifiers not followed by a noun) are sometimes treated as a ‘fused-head construction’ where a single word combines both the functions of determiner and head (cf. Chapter 5 in Huddleston and Pullum, 2002). Nevertheless, other analyses exist as well where the partitive construction is analysed as a noun phrase with an empty nominal head (e.g. Jackendoff, 1977).

(5:52) *Ja, wir ha'm halt die größte Brauereidichte der Welt in Bayern, deutschlandweit gibt's 1400 Brauereien, in Bayern gibt's rund 800 [/] davon.* (GO_INTERVIEW_013)

(5:53) *A range of companies, some [/] of them are based on technology, some [] are not.* (EO_INTERVIEW_007)

(5:54) *Neither [/] of the two theories, the idea of photons or the theory of relativity were accepted by very many people until fifteen or twenty years had gone by.* (EO_ACADEMIC_005)

Nominalised numerals are not annotated as ellipses. They are used as nouns and they can be part of proper names or fixed expressions (*'the Fantastic Four'* / *'the Dirty Dozen'*, *'der Rat der Vier / die Großen Vier'*). Inserting an additional noun from the textual or situational context would lead to a different analysis of the noun phrase. Nominalised German cardinal numbers referring to the numeral are feminine and inflected for number (*'zwei Vieren/r im Zeugnis'*, *'die Zehn als zweithöchste Karte'*). They cannot be followed by a noun and or interpreted a nominal ellipsis in any case.

5.1.4 Nominal ellipses after adjectives

Apart from ellipsis after numerals, both English and German use nominal ellipses after adjectives.

- (5:55) a) *But since 1992, the British economy has grown some 40 %, the German [] barely 10 %.* (EO_ESSAY_012)
- b) *Aber seit 1992 ist die britische Wirtschaft um rund 40 Prozent gewachsen, die deutsche [] nur um knapp zehn Prozent.* (GTRANS_ESSAY_012)⁸⁴

Here we encounter a major difference between the two languages. In English, nominal ellipsis after adjectives is not always possible where it can be used in German. In English, it is mainly restricted to ellipsis after comparative and superlative adjectives as those forms have a richer morphology and to some frequently used adjectives describing size, age, nationality, material or colour, especially if they express contrast to other noun phrases in the immediate context. To avoid explicitly mentioning or repeating a noun after adjectives, English can, or often even has to use the nominal substitute *'one[s]'* which is inflected for grammatical number and helps the reader or listener to clearly identify the preceding item as a noun modifier and the phrase as a noun phrase (5:56-57).

⁸⁴ In addition to nominal ellipsis, this example is a case of 'gapping'.

(5:56) *He was loth to abandon the rich brightness of the golden dame, or the lovely clarity of the silver one.* (EO_FICTION_002)

(5:57) *Not real animals though: fluffy ones playing games with coloured balls.* (EO_FICTION_008)

Conventional wisdom holds that ‘one’ in these examples is a regular common count noun meaning something like ‘instance thereof’, referring anaphorically to some type or class referred to in the discourse (Payne et al., 2013), but Kayne (2015) recently suggested a different analysis of ‘one(s)’ as a complex determiner whose relation to its antecedent is mediated by a silent noun. Anaphoric ‘one’ usually cannot be directly preceded by a numeral in standard English (* ‘two ones’) and it cannot substitute for mass nouns and uncountable nouns – either the noun has to be repeated or omitted (5:58)

(5:58) *I like sour milk better than fresh [].* (example from Bloomfield, 1933: 252)

The substitute ‘one’ as a proform has no exact equivalent in German where adjectives and other modifiers in an ellipsis remnant show strong morphological agreement in order to license the elided noun. This explains why nominal ellipsis can be used in more contexts in German (5:59-61).

(5:59) a) *Gerd drückte die Zigarette aus und zündete sich eine neue []*

an. (GO_FICTION_002)

b) Gerd crushed out the cigarette and lit a new one.

(ETRANS_FICTION_002)

(5:60) a) *Für den praktischen Gebrauch benötigt man größere Zahlen, solche mit mehr als hundert Dezimalstellen. Für das Kochrezept habe ich kleinere [] genommen. (GO_POPSCI_008)*

b) For practical application, large numbers are needed, numbers with more than a hundred digits. For our recipe I have chosen smaller ones. (ETRANS_POPSCI_008)

(5:61) a) *Sie sollten dann einige Menüpunkte sehen. Die oberen [] sind die Symbolleisten, welche Sie anzeigen lassen können.*

(GTRANS_INSTR_006)

b) You should see a few menu items. The top ones are the toolbars you may show. (EO_INSTR_006)

Using 'one-'substitution when it is optional or mandatory makes grammatical number of a phrase explicit in which the head noun has been omitted. It also helps to clarify that a phrase refers to a specific noun from the context or co-text and to distinguish adjectives used as modifiers from nominalised adjectives with generic or abstract reference (e.g. 'the best',

'the worst', denoting either a person or a thing that is better or worse than all others [5:62-63]).

(5:62) *The only way to be sure you will not regret a decision is by making the best possible one.* (EO_POPSCI_005)

(5:63) *The result is masters of the synonym who always replace a word or expression by the next worst one.* (ETRANS_FICTION_003)

In example (5:63), the translator of the text included in our corpus chose to use the phrase *'the next worst one'* to clarify that the phrase does not have a generic or abstract reference, but refers anaphorically to a specific noun. Leaving out the substitute in *'the next worst []'* may lead to a different interpretation of this noun phrase in the sense of *'the next worst thing'* / *'das Zweitschlimmste'*. It should be noted that both phrases *'the next worst one'* and *'the next worst'* in this example are problematic or marginal as they seem to combine features of the comparative and the superlative. (5:62) does not seem to be an ideal translation of the sentence from the corresponding original text in the GECCO corpus: *'Dabei kommen lauter Meister des Synonyms heraus, die einen Ausdruck immer durch den nächstschlimmeren [-> by one that is even worse] ersetzen* (GO_FICTION_003).

'One'-substitution also clearly distinguishes adjectives as modifiers from those that have undergone conversion and have acquired a specific

lexical meaning as lexicalised nouns (e.g. *the green*, denoting, for example, an area of grass in a village or on a golf course [5:64] or *the poor*, denoting people who are poor [5:65]).

(5:64) *This teaches you how to become a better golfer without having to step on the green!*⁸⁵ vs. *In the computer game, there's a green teleporter and a purple teleporter. Step on the green one.*

(5:65) *Wealth continues to be transferred to the rich nations from the poor ones* vs. *Wealth continues to be transferred to the rich from the poor.*

It is mainly after elements that can be sufficiently identified as noun modifiers in incomplete noun phrases and distinguished from nominalizations in their respective context and co-text that *one*-substitution is optional:

(5:66) *These are excellent biscuits. Can I have another ([]/one)?*

(5:67) *Jill's first semantics book was clearer than her second ([]/one).*

After determiners followed by inflectional superlatives (e.g. *the*

⁸⁵ Words that can belong to different word classes are often tagged incorrectly in the GECCo corpus, e.g. *the prospect of cricket on the village green*, EO_TOU_005, where *green* is tagged as an adjective. This makes it look like a potential nominal ellipsis in a query for adjectives in noun phrases that are not followed by a noun.

smallest, *the cheapest*') or periphrastic superlative constructions (e.g. *the most expensive*'), *one*' is left out in many cases:

(5:68) *It's the most expensive material there is. But, you know, people are buying the adobes and there is all the work to put them up. So it's the most expensive []. And yet, it's the cheapest [].*
(EO_INTERVIEW_011)

As this is an area of language contrast between English and German, it may involve certain translation difficulties. It is possible that translators overuse or underuse *one[s]*' to the detriment of stylistic variation or that they even sometimes use it incorrectly. Probably, when it is optional, *one*' tends to be inserted in translations, but is often left out in originals (5:70-71) to avoid its frequent repetition. In (5:69b) *one*' is optional and the phrase *the smallest one*' from this translation only occurs once in the whole GECCo corpus, while *the smallest []*' involving a nominal ellipsis occurs several times, and only in English original texts, particularly in one specific spoken text where the speaker talks frequently about *the smallest item / value*'. He uses nominal ellipsis in several cases, e.g. in the example given in (5:71), but he never uses the phrase *the smallest one*' for a change.

(5:69) a) *“Der Nagel” , sagte ich und schnitt den kleinsten [] ab.*

(GO_FICTION_002)

b) *“The nail,” I said, cutting off the smallest one.*

(ETTRANS_FICTION_002)

(5:70) *List your debts from the smallest [] to the biggest []. Start paying off the smallest [].*

(5:71) *First of all we check to see if the starting point element is the smallest [], because, if that is the smallest [], we don't need to make an exchange and we really oughta skip this.*

(EO_ACADEMIC_006)

5.1.5 Nominal ellipses after possessives and after classifier nouns

English possessive noun phrases take the possessive marker of apostrophe + *s* in singular noun phrases or *s* + apostrophe if it is a plural noun phrase. This is usually not analysed as a case inflection marker (Quirk et al., 1985: 328), but as a clitic. Some scholars, e.g. Zwicky (1987) analyse the English possessive as an ‘edge affix’. It can attach to single words but also to phrases and it does not necessarily attach to the head noun of possessive noun phrases. The English possessive marker leads to a slightly different structure of English possessive noun phrases compared to German possessive noun phrases (e.g. ‘*all the world’s poor*’ / ‘*die Armen der ganzen Welt*’ [EO_ESSAY_008]). Nominal ellipses are sometimes introduced by the possessive markers ‘*s* in English after different types of modifiers (5:72-74).

(5:72) *It’s actually got a GDP apparently that’s bigger than*

Namibia’s []. (EO_INTERVIEW_004)

(5:73) *The hips concealed beneath it cannot be appreciably bigger than*

her waist - not unless the latter is as narrow as an insect’s [].

(ETTRANS_FICTION_004)

(5:74) *When you’re writing a post, or replying to someone else’s, if it*

contains profane or inappropriate language, we will delete the

post. (EO_WEB_001)

Nominal ellipses can be introduced by a genitive marker, mainly after proper names, in German (5:75), but we do not find such occurrences in the corpus data.

(5:75) *Ich mag Antonias Auto, aber Martins [] ist besser.*

We do not annotate double genitives in ‘a [noun] of [noun]’s’-constructions as ellipsis. English double genitives have been explained with movement, rather than ellipsis, and have been claimed to be structurally identical to possessive partitives (e.g. ‘a friend of John’s’ = ‘one of John’s friends’, ‘two cars of John’s’ = ‘two of John’s cars’), cf. Barker (1995). The deletion of POSS-ing gerundive phrases after the possessive marker as in ‘Erik’s buying a car came as a surprise, and Bud’s [buying a car] came out of nowhere’ / ‘Mary’s writing a poem inspired John’s [writing a poem]’ has been claimed to be a type of verb phrase ellipsis (cf. LaCara, 2010). However, such mainly theoretically constructed examples are extremely rare in ‘real’ language use and do not occur in our data.

It has been claimed that possessive ‘pronouns’ also license nominal ellipses (Halliday and Hasan, 1976: 55) as in (5:76 and 5:77).

(5:76) *I was actually so happy with my dish that I forgot to try his [].*

(5:77) *Wenn Sie nur Ihre Änderungen beibehalten möchten, klicken Sie auf „Nur meine [] verwenden“.*

Halliday and Hasan state that the possessive form ‘*its*’ is rare in English nominal ellipsis but they provide an examples for its use ([5:78] cited from Halliday and Hasan, 1976: 46) and we do find a similar example in our corpus data (5:79).

(5:78) *You know that mouse you saw? Well that hole there must be its [].*

(5:79) *This is an opportunity for Russia to develop a new relationship with NATO that would advance not only our interests but also its [].* (EO_SPEECH_006)

There is some discussion on the exact status of possessives as pronouns, determiners or adjectives (Olsen, 1989). I generally agree with the assumption that a possessive can license nominal ellipsis in English and German. I have annotated cases in English and German where the insertion of a noun is not blocked by the inflectional suffix of a possessive pronoun. As this results in specific, and not all, forms of the inflection paradigm of possessive pronouns followed by nominal ellipsis, these cases have been marked as ‘problematic’ in the annotation.

A clear area of language contrast between English and German is the

possibility to omit nouns elliptically after English classifier nouns.

(5:80) *I prefer cotton shirts to nylon [].*

(5:81) *These shakes are really great. I have just tried the butter
pecan [], but I prefer the chocolate [].*

These English nouns behave like adjectives and refer to materials or substances for which there is no specific adjective such as ‘*wooden*’, ‘*golden*’, ‘*linen*’, ‘*brazen*’ (archaic). German in general has fewer zero-derivational relationships and the distinction between noun and adjective is clearer there. Nominal ellipsis after classifier nouns is very rare or even non-existent in our corpus registers.

5.1.6 Nominal ellipses after quantifiers

In both languages, nouns can be omitted after quantifiers and indefinite determiners. The following examples illustrate nominal ellipsis after quantifiers and indefinite determiners in English:

(5:82) *While Kim had lots of books, Pat had very few [].*

(5:83) *Each of these columns may have a different format, and each [] will contain 10,000 entries.*

(5:84) *This is a list of colleges and universities in metropolitan Boston. Some [] are located within Boston proper while some [] are located in neighboring cities and towns, but all [] are within the 128/95/1 loop.*

'Others' has a plural affix that blocks the possibility of inserting a noun in English. Only the singular form 'other' allows the insertion of a noun

(5:86).

(5:85) *Now, two words are used for this. One [] is exchange and the other [] is swap. (EO_ACADEMIC_006)*

(5:86) *Opting for one approach and neglecting the other [] risks failure.*

(ETRANS_SPEECH_003)⁸⁶

Equivalent structures in German also carry inflectional morphology, but can be analysed as ellipses in both singular and plural as nouns can be inserted after both forms:

(5:87) a) *We have to understand why some European economies are creating jobs, and others [/] are not.* (EO_ESSAY_003)

b) *Wir müssen herausfinden, warum manche Volkswirtschaften Europas Arbeitsplätze schaffen und andere [] nicht.*

(GTRANS_ESSAY_003)

Examples of nominal ellipses after quantifiers or indefinite determiners in German are (5:88-90).

(5:88) *Kim hatte viele Bücher, aber Pat hatte nur sehr wenige [].*

(5:89) *Es sind zwar Kisten dabei, die man nicht verwenden kann. Ich nehme trotzdem alle [].*

⁸⁶ This is an example from the corpus where a nominal ellipsis in the translation can have a ‘garden path effect’ on the reader. The tagger could not identify any finite verb in this sentence; both “risks” and “failure” have been tagged as nouns and the nominal ellipsis would not be found by a query for incomplete nominal phrases.

(5:90) *Zu Schottland gehören viele Inseln, einige [] sind mit dem Flugzeug erreichbar, aber die meisten [] sind von der Fähre abhängig.*

Quantifiers and indefinite determiners referring to people generally and not explicitly to nouns mentioned in the immediate co-text will not be treated as cohesive ellipses (e.g. ‘*many / some / all / viele / manche / alle / jeder*’). They function as independent pronouns. In a few cases, such cases are locally or globally ambiguous between a generic and an anaphoric interpretation (5:91).

(5:91) *First, the opponents of free trade claim to want to help poor countries but at the same time are reluctant to invest or trade with them. [...] Second, some [people / opponents of free trade?] place an emphasis on spreading democracy, but then ridicule the decisions of democratic governments. (EO_ESSAY_019)*

(5:92) *Irakische Professoren, die ihre Posten nicht dazu missbraucht haben, andere []⁸⁷ einzuschüchtern oder Verbrechen zu begehen, sollten ihre Arbeit unverzüglich wieder aufnehmen können. (GTRANS_SPEECH_004)*

⁸⁷ [?Personen / Iraker / Professoren]

The co-occurrence of numeral or quantifier and adjective followed by nominal ellipsis is possible in German while English mainly uses 'one' in similar cases (5:93-96).

(5:93) *Insgesamt sind es fünf Teiche, drei größere [] und zwei kleinere [].*

(5:94) *Es gibt nur wenige gute Projekte, oder besser gesagt, ich kenne nur wenige gute [].*

(5:95) *All the students took the exam but three lazy ones failed.*

(5:96) *There were no major problems, just a few small ones.*

5.1.7 Nominal ellipses vs. nominalisation

English and German de-adjectival nouns resulting from conversion can refer generically to abstract ideas or groups of people. Some of these de-adjectival nouns are ad hoc nominalisations and a few have become lexicalised and entered the dictionary. English nominalised adjectives mainly refer to specific groups of people with regard to age (*'the young'*, *'the elderly'*), social status (*'the unemployed'*, *'the rich'*), physical state (*'the blind'*, *'the sick'*), nationality (*'the Spanish'*) or features of character (*'the meek'*, *'the faint-hearted'*, *'the corrupt'*, *'the undecided'*). Some cases of generic group reference have inflectional or analytic comparative or superlative forms (*'the weakest'*, *'the most vulnerable'*, *'the brightest'*, *'the most committed'*, *'the more advanced'*) and some can be modified by *'very'* (*'the very weak'*). Usually these nominalisations take a definite determiner and they do not inflect for the genitive case or are morphologically marked for number in English, but take a plural verb. In a few cases they can be found with other determiners than *'the'* (*'these dead'* / *'our disabled'*, cf. Günther, 2013: 139). In rare cases, a plural affix can be added to a lexicalised nominalisation in English (*'the Greeks'*, *'some of the undecideds'*⁸⁸). Apart from generic group reference in certain cases, English nominalised adjectives can refer to some abstract concepts (*'the supernatural'*, *'the feminine'*, *'the delightful'*, *'the sacred'*, *'the shocking'*). Such cases of nominalised adjectives were regarded as

⁸⁸ cf. for instance: <https://www.washingtonpost.com/archive/politics/1998/12/16/the-undecideds/4712fe11-6d3b-4721-9585-521b7cea7e3b/> [last checked 25/07/2016]

exophoric ellipses by Halliday and Hasan (1976: 166). Huddleston (1988: 104) was another proponent of the view that such abstract and generic group reference constructions are different from de-adjectival conversion, although he admitted that it is difficult to clearly distinguish them from nominalisations. Tucker (2014) calls them ‘adjective-noun fused-head constructions’ and Günther (2013: 76) clearly prefers an ellipsis analysis of nominalised adjectives which she claims actually remain adjectives which modify a silent noun and can be described as ‘nounless noun phrases’ in both English and German. Günther (forthcoming) forcefully rejects the view that such adjectives undergo any conversion or nominalisation process. It is understandable why this view has become popular, but in my opinion, these cases should be clearly analysed as nouns, and not as ellipses. The absence of an overt number inflection marker in most English examples can be explained by the fact that they always refer to an abstract concept or have generic reference to a group (in contrast to other de-adjectival nouns that can refer to individuals or to groups, e.g. ‘*an intellectual*’ vs. ‘*several intellectuals*’). They will not be annotated as nominal ellipses.⁸⁹

A few English examples may seem ambiguous with regard to part of speech at first sight, but usually can be clarified from the context (5:97-98).

⁸⁹ Another structure that should not be analysed as exophoric or fossilised noun ellipsis are prepositions with adjectives as complements which are not followed by nouns (e.g. ‘*in particular*’, ‘*in general*’) functioning as adverbial phrases. It would sometimes be possible to add a noun to such a construction which is not grammatically incomplete without it (e.g. ‘*in general terms*’).

(5:97) *Two more nations - Afghanistan and Iraq - have thrown off the shackles of tyranny and are joining the ranks of the free [nations].* (EO_SPEECH_013)

(5:98) *In the past 4 years alone, more than 110 million human beings across the world have joined the ranks⁹⁰ of the free [/].*

Nominalised adjectives in the form of zero derivation or conversion are more frequently used in languages with adjective agreement morphology. Ancient Greek, for instance, used nominalised adjectives and quantifiers (e.g. οἱ πολλοί, lit.: ‘the many’ referring to the masses, the people, οἱ ὀλίγοι lit.: ‘the few’ in the sense of ‘the aristocracy’). In languages such as Greek, French or German, nouns and adjectives carry inflectional endings and therefore adjectives in these languages can be converted into nouns in more contexts than in English, referring to a single person, groups of people or abstract concepts. Additionally, there are numerous German participles that can be used as adjectives or undergo nominalisation (‘*ein Verletzter*’, ‘*das Getane*’, ‘*etwas Gekochtes*’, ‘*etwas Leuchtendes*’, ‘*der Schwimmende*’ etc.). Nominalised adjectives start with a capital letter in German because of their noun-like status (5:99-100):

(5:99) *Die Einheimischen* [/] (=group of people) *waren sehr freundlich.*

⁹⁰⁹⁰⁹⁰ The word ‘*ranks*’ already implies that the following specification refers to people who belong to a certain group, e.g. *the ranks of the unemployed*, therefore we do not have to add an additional word such as ‘*people*’ after the nominalised adjective cf. <https://www.collinsdictionary.com/dictionary/english/ranks> [last checked 01/06/2016]

vs. *Fremde Baumarten sind weniger gut geeignet als einheimische [Baumarten].*

(5:100) *Er bestellte ein Helles [/].*⁹¹ vs.

In Bayern gibt's mehr Landbiere und helle [Biere].

(GO_INTERVIEW_013)

In several corpus examples the spelling is wrong if the difference between nominalisation and nominal ellipsis has not been recognised. Some nouns originally derived from other parts of speech such as adjectives or numerals wrongly begin with small letters in the German corpus data (5:101-104), particularly in transcriptions of spoken language, in translated texts or in texts from registers that were not extensively edited or proofread.

(5:101) *So dürfen auch Sie das gleiche erwarten.* (GTRANS_TOU_10)

(5:102) *Ich war gleich von Anfang an auf dem laufenden.*

(GTRANS_FICTION_001)

(5:103) *Obwohl jeder im stillen genau dies befürchtet.*

(GTRANS_FICTION_005)

⁹¹ <http://www.duden.de/rechtschreibung/Helles> [last checked 19/02/2016]

(5:104) *Ein riesiges Landhaus und Gut, das für 200 Jahre der Familie York gehörte und einen unvergesslichen Eindruck davon liefert, wie das Leben war, nicht nur für die sehr reichen, sondern auch für die weniger glücklichen, die ihnen das Leben so angenehm machten.* (GTRANS_TOU_009)

Capital letters have wrongly been used at the beginning of some German adjectives or numerals in ellipsis remnants (5:105-106). Wrong capitalization and spelling errors result in tagging mistakes that would make an automatic identification of such nominal ellipses impossible.

(5:105) *Beeindruckend sind die weit über 1000 Linien, die die Hochebene überziehen. Die Längste erstreckt sich über ca. 9 Kilometer.* (GO_WEB_004)

(5:106) *Bei den Älteren stimmt das nicht mehr.* (GO_INTERVIEW_012, referring anaphorically to children that are older than other children from the group and not to elderly people in general ('*die Älteren*').

Some texts show inconsistencies with alternations between different spellings in the same document (5:107-108).

(5:107) *Na ja, erklärt die zweite, rechnen Sie doch um. [...] Die Dritte,*

*eine dürre Frau, die bisher wenig gesprochen hat, bemerkt jetzt
mit stillem Triumph [...] (GO_FICTION_010)*

*(5:108) In manchen Ländern werden vierjährige an den Teppichwebstuhl
gebunden, in anderen Länder helfen Siebzehnjährige auf dem
Bauernhof der Eltern. (GTRANS_WEB_012)*

The phrases ‘die (Z/z)weieinhalbjährigen’ and ‘die (K/k)leinen’ / ‘die (G/g)roßen’ in the following German examples (5:109-110) are ambiguous in spoken language between an interpretation as a nominalisation (as in 5:111-112) and a modifier with nominal ellipsis referring back to a noun from a previous sentence (5:113-114).

*(5:109) Die Zweieinhalbjährigen, die sind also wirklich schon dabei,
miteinander zu spielen. (GO_INTERVIEW_012)*

*(5:110) Kannst du mal was erzählen so über den Tagesablauf jetzt in
deiner Kindergruppe, auch vor allen Dingen darüber, dass es ja
jetzt die ganz kleinen Kinder sind [...]? [...] Ist halt dann
schwierig mit den Kleinen, [...] die müssen wir dann einfach zur
Seite nehmen, damit die Großen halt auch ihren Freiraum haben.
(GO_INTERVIEW_012)*

(5:111) Auch mit einem Zweijährigen [=Junge im Alter von zwei Jahren]
sind Abmachungen möglich.

(5:112) Der Große kommt bald zur Schule; der Kleine erst in zwei
Jahren.

(5:113) Man muss mit Kindern sprechen. Auch mit einem zweijährigen
[Kind] sind Abmachungen möglich.

(5:114) Zwei Jungen spielen Fußball. Der kleine [Junge] jagt dem
großen [Junge] dauernd den Ball ab.

A few spelling inconsistencies between different texts can be explained by recent changes in the German language due to a spelling reform that affected capitalization rules. Particularly, the decision between upper or lower case letter with certain quantifiers and indefinite pronouns is now optional and has led to some confusion, inconsistencies or common misspellings in the corpus data due to wrong analogies.⁹²

In written texts and in the transcriptions of spoken texts, the authors had

⁹² According to the orthographic rule §77 in the Duden, words such as ‘viel’ / ‘wenig’ / ‘[der] eine’ / ‘[der] andere’ may start with a capital letter to emphasise their ‘noun-like’ role (‘Das Lob der V/vielen (= der breiten Masse) war ihr nicht wichtig.’ / ‘Auf der Suche nach dem A/anderen (= nach einer neuen Welt) sein.’ / ‘Die E/einen sahen zu, die A/anderen halfen mit.’ / ‘Die M/meisten blieben zu Hause.’ ‘Es gab viele, die nicht mitmachen wollten.’ / ‘Den wenigsten war das bekannt’, cf. <http://www.duden.de/sprachwissen/rechtschreibregeln/Gro%C3%9F-%20und%20Kleinschreibung> [last checked 12/02/2016] and the GECCo project annotation guidelines (Menzel, 2014b) for a more detailed discussion and more examples).

to make a decision and mark the part of speech clearly, however it is not always correct in the corpus data in GECCo. Where the adjective should correctly begin with a small letter in the German corpus data as it is used in an elliptical phrase and not as a nominalization, it is annotated as an ellipsis, particularly in the case of anaphoric reference to the immediate co-text. In the English corpus data, there are fewer ambiguities as a result of spelling inconsistencies. One particular aspect that leads to tagging mistakes in English and that makes automatic queries for incomplete noun phrases difficult is the frequency of conversions and zero-derivation as an English word formation process. In the following example ‘*black*’ and ‘*white*’ are used as nouns but have been tagged as adjectives. As they follow a determiner, the tagged structure wrongly suggests an incomplete noun phrase:

(5:115) *There's not really a black and white.* (EO_FORUM_005, tagged as: a_DT black_JJ and_CC white_JJ)⁹³

Sometimes an English plural suffix can clearly mark a lexicalised deadjectival noun blocking the insertion of an additional noun (e.g. ‘*5-year olds*’ (EO_WEB_010), ‘*empties*’⁹⁴, ‘*classifieds*’⁹⁵). A few English phrases

⁹³ The English tag set in GECCo is based on the Penn Treebank Tagging Guidelines (<http://www.clips.ua.ac.be/pages/mbsp-tags>) and the German tag set is based on the Stuttgart-Tübingen-TagSet STTS (<http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/German> [last checked 08/02/2016])

⁹⁴ Colloq.: bottles or glasses that are empty, <http://www.ldoceonline.com/dictionary/empties> [last checked 12/02/2016]

include a nominalised use of *one[s]* as in *'the little ones'* (= children, ETrans_TOU_019) and *'loved ones'* (= family members, EO_SPEECH_005). As English usually does not mark nouns by a capital letter, information about the part of speech can generally be taken from reliable dictionaries such as the OED or the Longman Dictionary. There are minor differences between English and German with regard to part of speech analysis so that a few very similar structures would count as nominal ellipsis in one language, but not in the other, e.g. *'the former'* and *'the latter'*⁹⁶ are listed as nouns, therefore they will not be analysed as incomplete noun phrases in English.

⁹⁵ Orig. U.S.: a small advertisement placed in a classified section of advertisements in a newspaper, chiefly in pl. (usu. with definite article), cf. <http://www.oed.com/view/Entry/33900>, sense B [last checked 12/02/2016]

⁹⁶ http://www.ldoceonline.com/dictionary/former_2/ / [last checked 12/02/2016]
http://www.ldoceonline.com/dictionary/latter_1/ [last checked 12/02/2016]

5.1.8 Nominal ellipses vs. Right Node Raising and ‘ellipsis’ within the noun phrase

There are several cases of Right Node Raising of nouns or ‘Tilgung’ within noun phrases in the context of coordination of several noun phrases or coordination within noun phrases that have been called ‘ellipsis’ by various authors. In my opinion, some cases that may be seen as cataphoric nominal ellipses are better analysed as modifiers of the same heads or Right Node Raising (RNR) structures. Right Node Raising, a term originally attributed to Postal (1974: 125) – although the phenomenon had also been described earlier by Ross (1967) – is a construction whose analysis has mainly evolved along with generative theory (cf. Hartmann, 2000: 53). RNR has been analysed as across-the-board extraposition or movement, which moves a constituent out of all the conjuncts of a coordinate structure (Ross, 1967: 107, Abels, 2004). It has also been called backward ellipsis or backward conjunction reduction (Ross, 1967), right periphery ellipsis (Höhle, 1991) or constituent sharing in coordination / shared constituent coordination (Radford, 1988) where a shared argument surfaces at the right periphery of a coordinate structure. In German it is often called ‘Linkstilgung’. Head nouns can undergo Right Node Raising as in (5:116-118) and in (5:119a).

(5:116) *We do want not just economic ___⁹⁷ but also social stability and development.*

⁹⁷ In the examples here and below, RNR is indicated by ‘___’.

(5:117) *both in front of the blue __ and behind the white house* (example taken from Haspelmath, 2007: 41)

(5:118) *Anfangs in jedem zweiten __, seit 1988 in jedem Jahr nutzt seither ein Gastland das bedeutendste Treffen der internationalen Büchergemeinde, sich mit seiner Kultur und Literatur vorzustellen.* (GO_WEB_006)

(5:119) a) *The impression was given that the necessary transfers of public funds from western __ to eastern Germany for restructuring and development could easily be financed out of “petty cash”.*
(ETrans_ESSAY_003)

b) *Andererseits ließen sich die zum Um- und Aufbau erforderlichen öffentlichen Transfers von West- nach Ostdeutschland quasi aus der „Portokasse“ finanzieren.*
(GO_ESSAY_003)

In the German original of (5:119), the German equivalents of ‘*western Germany*’ and ‘*eastern Germany*’ are compounds that are written as single words and therefore a ‘suspensive’ or ‘hanging’ hyphen (Ergänzungsbindestrich) marks the dropping (‘Einsparung’) of the common element in a series of compound words.⁹⁸ This type of

⁹⁸ There are various patterns of suspensive hyphenation in English and German. With compounds or affixed forms in a series containing a common element, the hyphen is usually retained with the first element of all but the last (‘*Grün- und Buntspechte*’, ‘*Um-*

‘Einsparung’ or ‘Tilgung’ within noun phrases has also sometimes been called ‘ellipsis’ or ‘elliptical coordination’. Additionally Buyko et al. (2007), for instance, very broadly subsumes other noun phrase structures with several modifiers or several head nouns under the term of ellipsis, e.g. *‘human and mouse cells’*, or *‘human genes and cells’*. This represents an extreme view that would lead to an ellipsis analysis of almost all non-sentential types of coordination. This view was popularised by Quirk and Greenbaum (1973b: 268) who saw examples such as *‘old and young men’* as ellipsis of the head where the full form would be *‘old men and young men’* as the men in this example are not both old and young at the same time. If we counted this as ellipsis, we would have to annotate an enormous number of this type as nominal ellipsis which would unreasonably skew the results. We would find quite a lot of these structures in our corpus data and examples such as *‘small, medium-sized and large companies, organisations and institutions’* would have an incredibly long underlying full form of the phrase. Therefore we neither include these cases of noun phrases nor noun phrases with suspensive hyphenation in series of compounds or affixed forms in the analysis of ellipsis, and certainly not in the analysis of cohesive devices.

und Aufbau’, *‘in- and output’*, *‘six-, eight-, or tenfold’*, *‘a shift from hard- to software’*). The different elements are usually connected by *‘and’*, *‘or’* or *‘to’* and their German equivalents. Other forms omit shared elements from the beginning of a word or from both sides (*‘Grünspechte und -finken’*, *‘an NBC-owned and -operated station’*, *‘Warenein- und -ausgang’*). Uses such as *‘applied and sociolinguistics’* or *‘Grün- und sonstige Spechte’* are frowned upon in both English and German.

Due to its similarity to nominal ellipsis, Right Node Raising can be a translation strategy for ellipsis, but it leads to a different focus structure (5:120).

(5:120) a) *If we analyse our problems, we will see that Europe needs more competition, not less [].* (EO_ESSAY_012)

b) *Wenn es seine Probleme genau analysiert, sieht es nämlich, dass Europa nicht weniger __, sondern mehr Wettbewerb braucht.* (GTrans_ESSAY_012)

Particularly in the English corpus texts, we also find various examples of right node raising structures with stranded prepositions if two constituents – nouns or verbs – are completed with complements that need different prepositions or if only the first constituents takes a complement with a preposition:

(5:121) *We have strengthened our commitment to and support for the principle of joining the euro.* (EO_ESSAY_005)

(5:122) *The Clinton administration has listened to and agrees with some of the concerns related to free trade.* (EO_ESSAY_019)

(5:123) *? This guarantee statement is in addition to and in no way prejudices your statutory rights.* (EO_INSTR_008)

RNR is possible for nouns in constituent coordination, but also for verbs in coordinated clause complexes where it shares some similarities with gapping (Chapter 5.2.2). At first sight, RNR in (5:124) may not seem very different from anaphoric nominal ellipsis in (5:125). Additionally, in these cases, RNR of the noun and nominal ellipsis has been combined with ellipsis of the verb in a gapping structure.

(5:124) *Der eine Polizist bewachte den rechten___, der andere [] den linken Eingang.*

(5:125) *Der eine Polizist bewachte den rechten Eingang, der andere [] den linken [].*

An important difference between RNR and anaphoric ellipsis is that in RNR, the first part of the structure cannot be separated from its second part. In the example of the anaphoric nominal ellipsis, the second part is much more independent of the rest of the sentence. A sentence split or a turn-taking between both parts is possible (*A: Der eine Polizist bewachte den rechten Eingang. B: Der andere den linken [].*) As the concept of RNR is not entirely clear from the literature, such structures will not be annotated as ellipses in our corpus. In my opinion, RNR is best analysed as a type of constituent sharing in coordination where a shared argument surfaces at the right periphery of a coordinate structure and there are some clear differences to anaphoric ellipses. Right node raising does not seem to

allow grammatical mismatches and can involve elements below the word level.

Moreover, we may find a few discontinuous noun phrases in German whose constituting elements are split in a topicalisation construction, particularly in spoken language (e.g. *'Lieder haben die Musiker keine schönen gespielt'*, *'Bücher sieht Anna drei'*⁹⁹). Similar structures have been discussed with regard to English as split noun phrases ('As for books, Anna can see three. '), quantifier floating or quantifier stranding, e.g. for the quantifiers *'all'*, *'both'* and *'each'* (e.g. de Mönnink, 2000: 144). Although these structures do not count as omissions, it would probably be difficult to systematically rule out such cases of modifiers not followed by a noun in an automatic annotation procedure.

Our annotated and extracted results on nominal ellipses in the GECCo corpus have been sorted in lists with additional information on the specific subtype of ellipsis so that it is possible to filter the results by additional criteria, e.g. to look specifically at nominal ellipsis after cardinal and ordinal numbers, after *'one'* as a particular case, after adjectives in their base form, their comparative or superlative form, quantifiers, possessive markers or possessive determiners or classifier nouns. Cases that may seem to be on the borderline to other cohesive devices in the opinion of some or that may raise doubt of whether they should be analysed as ellipses at all – for instance, if there are different opinions in the literature

⁹⁹ cf. also Ott, 2012 for a detailed analysis of split topicalisation and quantifier float in German

– have been marked as problematic in the annotation. Therefore, if one wants to narrow the scope of the ellipsis definition, it is possible to consider only those cases for the calculations that in the opinion of the reader are typical, perfectly clear cases.

5.2 Verbal ellipses

5.2.1 The verb phrase in English and German

In this chapter, verbal ellipsis and its subtypes will be described and illustrated with examples. Before explaining the structures that fall under the category of verbal ellipsis in the annotation scheme, a precise definition of what we mean by the term ‘verbal ellipsis’ will be presented. Then, a brief overview of the internal structure of the verb phrase in English and German will be given in this section to compare the possible structures of verbal ellipsis remnants and the syntactic configurations that allow verbal ellipsis in these languages. Verbal ellipsis, like nominal ellipsis, is a concept that we can use for a cross-linguistic analysis of the English-German language pair. Similarities between English and German verb phrases as well as language-specific differences will be pointed out.

Particularly among generative syntacticians, verbal ellipsis has been discussed extensively in the theoretical literature from the 1970s onwards, e.g. by Ross, 1969 and Hankamer and Sag, 1976; Sag, 1976, and it continues to be a dominating subfield of the theoretical literature on ellipsis. While ‘verbal ellipsis’ is the SFL term that Halliday and Hasan (1976) use to refer to omissions of verbs, frequently the term ‘verb phrase ellipsis’ or ‘VP-ellipsis’ (in analogy with the frequently used term NP-ellipsis) is used in the literature. This usually does not refer to the omission of an entire verb phrase but is used interchangeably in publications with the term of the ‘elliptical VP’ (e.g. in Hardt, 1993). VP-ellipsis has been

described as an anaphoric phenomenon in which a missing predicate is linked to an antecedent in the surrounding discourse (Johnson, 2001); Quirk et al. (1985) used the term ‘predication ellipsis’. Its definition has sometimes been restricted to auxiliary verbs followed by ellipsis of the main verb (Hardt, 1992b¹⁰⁰). I define verbal ellipsis as an omission within a verb phrase or – to use the SFL term – within a verbal group. In both English and German, specific elements or a combination of several elements can be omitted in verbal ellipsis – in contrast to nominal ellipsis where the head noun as a specific grammatically necessary element of the noun phrase is always left out.

It has been claimed that German does not have VP-ellipsis at all (Beck and Gergel, 2014: 257). More precisely, German does not have VP-ellipsis of the English type, but it does have some possibilities for omitting elements from verb phrases that are similar in English and German (e.g. gapping, lexical verb ellipsis after modal verbs) while some subtypes of verbal ellipsis only seem to exist in English, but not in standard modern German (e.g. lexical verb ellipsis after stranded auxiliary verbs at the end of a clause, pseudogapping). Despite some differences with regard to the internal structure of English and German verb phrases, verbal ellipsis can involve the omission of an auxiliary or modal verb, a lexical verb or a combination of several verbal constituents from the verbal group in both languages. These cases are annotated as verbal ellipsis in our corpus data.

¹⁰⁰ To be precise, Hardt stated that VP ellipsis can be ‘defined by the presence of an auxiliary verb, but no VP’ which is slightly strange as the auxiliary verb is a VP, albeit an elliptical one.

Verbal ellipses that contribute to the cohesiveness of a text fall into two categories in Halliday and Hasan (1976), ‘operator ellipsis’ and ‘lexical ellipsis’. I avoid the term ‘lexical ellipsis’ for ellipsis of the lexical verb in order to forestall terminological confusion with the similar-sounding German term ‘lexikalische Ellipse’ (Ågel, 1991) which is used for conventionalised or lexicalised ellipsis. Sag’s (1976: 53) terminological suggestion for lexical verb ellipsis was ‘post-auxiliary ellipsis’ and Levin (1979) called it ‘main-verb ellipsis’. I will refer to the omission of lexical verbs by the term ‘lexical verb ellipsis’.

Although Halliday and Hasan described different types of omissions in verb phrases as types of potentially cohesive ellipses, we assume that cohesive verbal ellipsis prototypically involves the omission of the lexical verb as the head of the phrase. Scanning the linguistic literature for information on the structure of noun and verb phrases, one may get the impression that the headedness of the verb phrase is not addressed as frequently as the headedness of the noun phrase. Nevertheless, where grammarians made claims about head elements of English or German verb phrases, they usually stated that a core verb phrase consisting of a lexical verb, which can have modal and auxiliary verbs as closed class words as premodifiers or supplements as well as predicative complements, has the main verb as its immediate head. At the same time, the main verb is the ultimate head of the extended verb phrase (EVP) (e.g. Huddleston, 1984: 177, Hjulmand and Schwarz, 2012: 191, Fabricius-Hansen, 2010: 174). There are also authors who claim that, instead of the main verb, the

finite verb is the head of the verb phrase even if it is an auxiliary (Wachtel, 2004: 151, cf. also the critical discussion of this view in Huddleston, 1984: 143).

If the antecedent verb phrase from a previous clause or sentence included pre- and postverbal modifiers or complements, these can also be omissible. Modal and auxiliary verbs in long verb clusters can be omitted together with the lexical verb, but at least the finite verb has to remain as a remnant structure. Therefore this type of ellipsis has been called ellipsis ‘from the right’. It is only under a few specific circumstances that operators can be omitted while the lexical verb remains as a remnant structure. Auxiliary drop or the omission of function words in general as a reduction strategy without any additional omission of content words usually does not contribute to increased cohesiveness of a text (cf. Chapter 5.2.3 on operator ellipsis). Ellipses that function as cohesive devices are most typically linked to content words as their antecedents in a text.

The English and the German verb phrases function as sentential predicates expressing something about the subject. They consist of obligatory elements and can also contain additional optional elements. There are a few differences with regard to the internal structure of the verb phrase in English and German that are more striking than the subtle structural differences of the English and German noun phrase. The different inflection paradigms, the syntactic behaviour of auxiliary, modal and lexical verbs as well as the order of constituents in verbal clauses and the position of the verb in different clause types are well described in

monolingual standard grammars and will not be repeated here in great detail. We will mainly point out certain specific differences between English and German that have an influence on the possibilities for verbal ellipsis.

On the one hand, English and German use simple verb forms consisting of a finite main verb. On the other hand, both languages use complex or periphrastic verb forms consisting of at least one finite auxiliary and a lexical verb. Doherty (1999: 124) stated that German has a right-peripheral verb phrase while English has a left-peripheral verb phrase. Therefore, the grammatically determined focus can be expected more to the left in the English verb phrase and more to the right in the German verb phrase.¹⁰¹

An overview on some cross-linguistic differences between the English and the German verb phrase is given by König and Gast in their textbook on English-German contrasts that compares the systems of tense, aspect and voice of both languages as well as the position of the verb phrase in different sentence types (König and Gast, 2012: Chapters 5, 8 and 10). König and Gast briefly mention the internal structure of the above-discussed noun phrase which according to them is ‘by and large identical’

¹⁰¹ The inflectional paradigm of verbs in German is richer than in English. German strong verbs with stem alternation can have up to three stem forms and more than twenty different finite forms and several non-finite word forms for the categories person, number, tense, mood, imperative, infinitive and participles. The number of different forms in the inflectional paradigm of English lexical verbs ranges from eight for the verb ‘be’ (*be, am, are, is, was, were, being, been*), to only three for some verbs (e.g. ‘put’ with *put, puts, putting*). The different morpho-syntactic categories of the verbal inflectional paradigm in English are: past, non-past (3rd person sg., non-3rd person sg.), infinitive/subjunctive/imperative, present participle and past participle. English modal verbs, at least the core modal verbs, behave as auxiliary verbs do and have impoverished inflectional paradigms lacking participles and 3rd-person singular forms or sometimes even past tense forms (cf. Bauer et al., 2013: 61f.).

in English and German (2012: 208f.). In their remarks on verb phrases from a comparative perspective, they did not make any particular statement on the internal structure of the verb phrase or the distribution of constituents within the verb phrase, but rather focused on the description of rules for the position of the verb phrase within the clause. Neither did they compare any omission possibilities in phrases or clauses in general. A detailed analysis of verb phrase complement omission possibilities in English and German is given in Fischer's work on verb valency (1997) where he describes the range of optional and obligatory complements of verbs, but not the omission of verbs themselves. Winkler (2005, Chapter 3) describes some aspects of VP-anaphora and ellipsis of the VP in English and German.

As a general working hypothesis it appears reasonable to assume that verbal ellipsis is more frequent in English than in German as some specific subtypes of verbal ellipsis occur more frequently or virtually only in English. Additionally, English has more and longer verb phrases with more possibilities than German to leave out verb phrase elements. We can put forward the hypothesis that there are generally more and longer noun phrases in German which allow omissions and that there are more and longer verb phrases and more clauses in English where certain elements can be omitted. It is widely believed that English prefers a discourse style with more verb phrases (Verbalstil) than German which tends to use fewer verbs and more and longer noun phrases (Nominalstil). For instance, *'the approach we have charted'* has become *'der gewählte Ansatz'* in a

translation from English into German (EO/GTRANS_ESSAY_001), or *'the regime the EU is proposing'* corresponds to *'das von der EU vorgeschlagene Regime'* (EO/GTRANS_SPEECH_007) where we find finite verbs in English contact relative clauses¹⁰² adding further information about the head noun and participles used as adjectives in the German texts. In German, most present and past participles can be used as attributive adjectives and be expanded leftwards by adding more material to a phrase whereas English would often put this information into a separate clause, for instance a relative clause placed after the noun (cf. Chapter 5.1.1 on the differences between the noun phrase in English and German).

Moreover, modality is sometimes expressed by adverbs or modal particles in German where English tends to use modal verbs such as *'may'* and *'might'*, as in *'the site might have been tampered with'* which has been translated as *'die Site wurde möglicherweise manipuliert'* (EO/GTRANS_INSTR_009). *'The complexity of prokaryotes may have been limited'* became *'die Komplexität der Prokaryoten war möglicherweise limitiert'* (EO/GTRANS_POPSCI_004) and *'the things she should have been doing'* corresponds to *'die Sachen, bei denen sie eigentlich mitmachen wollte'* (EO/GTRANS_FICTION_001).

This is one aspect that typically leads to fewer and shorter verb phrases and fewer clauses with finite verbs in German. Another difference between

¹⁰² 'Contact' relative clause: a relative clause in which no relative pronoun intervenes between the head noun and the relative clause.

the English and the German verbal system in this regard concerns the frequent use of the German simple present with future meaning where English has to use a periphrastic future construction with an additional auxiliary in most cases. Furthermore, English verb phrases in conditional sentences, apart from those in dependent if-clauses, often occur with a modal verb such as ‘*would*’, ‘*should*’, or ‘*might*’ where German may either equally chose a modal verb or directly mark lexical and auxiliary verbs for modality with the subjunctive form *Konjunktiv II* as in (5:126) and in (5:127). On the whole, this contributes to simpler verb phrase structures in German with fewer possibilities for ellipsis within the verbal group and to more complex verb phrases in English with more omission possibilities for verbs where modal or auxiliary verbs can potentially be left as verbal remnants.

(5:126) a) *Nor has Microsoft demonstrated: first, that the use by its competitors of the information disclosed would lead to its ‘dilution’; second, that the fact that the competing products would remain in the distribution channel after the decision has been annulled would constitute serious and irreparable damage; third, that Microsoft’s competitors might ‘clone’ its products; fourth, that Microsoft would be required to make a fundamental change in its business policy; and fifth, that the decision would cause an irreversible development on the market.*

b) *Auch hat Microsoft nicht nachgewiesen, dass erstens die*

Nutzung der preisgegebenen Informationen durch ihre Konkurrenten darauf hinausliefe, dass die Information "verschwände", dass zweitens der Verbleib von Konkurrenzprodukten in den Vertriebskanälen nach einer möglichen Nichtigklärung der angefochtenen Entscheidung einen schweren und irreparablen Schaden darstellte, dass drittens die Konkurrenten von Microsoft ihre Produkte "klonen" könnten, dass viertens Microsoft ihre Handelspolitik fundamental umstellen müsste und dass fünftens die Entscheidung zu einer nicht wieder rückgängig zu machenden Entwicklung des Marktes führte.¹⁰³

- (5:127) a) *Just last autumn, who would have thought that Germany would be able to introduce the toughest social reform in its history without serious resistance? (ETRANS_ESSAY_002)*
- b) *Wer hätte im Herbst geglaubt, Deutschland könnte die härteste Sozialreform seiner Geschichte ohne ernsthaften Widerstand einführen? (GO_ESSAY_002)*

Furthermore, the progressive auxiliary 'be' in progressive constructions as an aspect marker is part of the verbal inflection system of English, while verbs in standard German are not explicitly marked for progressive aspect.

¹⁰³ example taken from press release by the European Commission on a court decision: http://europa.eu/rapid/press-release_CJE-04-103_en.htm?locale=en / http://europa.eu/rapid/press-release_CJE-04-103_de.htm?locale=en [last checked 09/03/2016]

The German progressive can be marked by adverbs. An additional reason that explains why English tends to have more constructions in which verbs cluster is the periphrastic use of the auxiliary 'do' as a question marker, a negative statement marker, emphatic particle or as an auxiliary in inversion contexts. The use of 'tun' ('do') as a dummy auxiliary in a periphrasis is very restricted to a few constructions; apart from those it is generally considered bad style in German. English progressive verb forms and the use of the periphrastic use of the auxiliary 'do' generally contribute to longer verb clusters in English. In certain English verbal ellipsis contexts, the lexical verb is omitted after the operator 'do' or the progressive auxiliary 'be' which have no exactly equivalent structures in German. German would only have a simple verb phrase consisting of a main verb in similar contexts.

On the other hand, German, like all Germanic languages, can also have relatively long verb clusters (Wurmbrand, 2006). German, for instance, uses the present perfect, where a finite form of an auxiliary is combined with the participle of the main verb, to talk about different types of past events while this complex verb form usually cannot have a specific past reference in English and there is a tendency to use the simple past instead (Klein and Vater, 1998). German also seems to allow more constructions containing clusters of double modals that are grammatically acceptable that have simpler equivalents in standard English (cf. Siemund, 2004: 23).

Omissions of verbs and omissions of entire verb phrases have been examined in great detail in monolingual studies by various authors,

particularly with regard to English, German and Dutch, and some other languages, if the respective structure exists in those languages (Ross, 1969; Bouton, 1970; Sag, 1976; Webber, 1978; Johnson, 1996, 2001; Lasnik, 2010; Merchant, 2001, Scheffler, 2005; Van Craenenbroeck, 2010 etc.). Contrastive analyses of verbal ellipsis usually focus on very specific structures, cf., for instance, Repp (2009b) who discusses negation in gapping in English and German in great detail. Winkler (2005) has a chapter on ellipsis of the verb phrase in English and German where she discusses the syntax of VP-ellipsis in English and of comparable instantiations in German. She also mainly describes gapping and particularly stripping as a subtype of gapping where only one contrast between two conjuncts is left with *'too / auch'* or *'not / nicht'* as polarity elements as in (5:128) and in (5:129).¹⁰⁴ Both languages can have an ellipsis of the verb phrase and its complement, but English can also have an ellipsis of the lexical verb after the operator which is not always possible in German. Example (5:128), for instance, demonstrates that omissions of lexical verbs after modals or auxiliaries are permitted in more contexts in English.

(5:128) a) *Leon can solve the task and Peter can, too.*

b) *Leon kann die Aufgabe lösen und Peter auch.*

¹⁰⁴ If we compare English and German gapping and sluicing structures to similar constructions in Dutch, we note that their structure more strongly resembles German syntactical patterns, e.g. *'Jan schrijft heel netjes maar Frits ijselijk slecht.'* / *'Ik leer hem Frans en hij mij Duits.'* / *'Jij schijnt het te gelooven, ik niet.'* (examples from Van Ginneken, J., 1910 – all these examples can be translated word for word into German: *'Jan schreibt sehr schön, aber Fritz furchtbar schlecht.'* / *'Ich lehre ihn Französisch und er mich Deutsch.'* / *'Sie scheinen es zu glauben, ich nicht.'*)

(5:129) a) *Leon can solve the task but not Peter.*

b) *Leon kann die Aufgabe lösen, aber Peter nicht.*

(5:130) is an ellipsis of the lexical verb after a negated form of a modal verb in English. The negated modal verb is part of the ellipsis remnant in English while the modal and the lexical verb are left out in German, which demonstrates the close relationship between the concept of verbal ellipsis and clausal ellipsis (the omission of a constituent from a clause, e.g. an entire verbal group).¹⁰⁵

(5:130) a) *We will find out which children can read and which cannot [read].*

b) *Wir werden herausfinden, welche Kinder lesen können und welche nicht [lesen können].*

As a general claim, it can be suggested that due to some structural differences between German and English verb phrases, different parts of the verb phrase can be omitted in similar contexts.

¹⁰⁵ Detailed information on the interplay of modals and ellipsis in English can be found in Gergel (2005).

5.2.2 Cohesive and non-cohesive cases of verbal ellipses

Verbal ellipsis is usually an endophoric phenomenon. Exophoric and deictic expressions in general are very limited in written corpus data as these contexts usually do not promote nonverbal reference and most references are usually contextualised within the text. In spoken data, exophoric interpretations in reference to the speech situation often depend on the use of gestures and other non-verbal cues and we do not find cases in our corpus data where a verbal ellipsis is used in a deictic function to point to the situational context without any relation to a co-textual antecedent. Exophoric verbal ellipses have not received very much attention in literature, but a few English structures involving ellipsis of the lexical verb after auxiliaries and modals have been described as being exophoric, e.g.: *'Shall we []?'* / *'May I []?'* / *'Don't []!'* / *'You shouldn't have []!'* cf. Schachter (1977). Only some of these structures have a German equivalent involving a verbal ellipsis (e.g. *'Sollen wir []?'* / *'Können wir []?'* / *'Darf ich []?'* /) and they seem to be restricted to spoken face-to-face conversation. The range of such constructions is very limited in English and German and they are often semantically conventionalised or virtually lexicalised. A sentence such as *'Shall we?'* can often be explained as a conventionalised courteous invitation to dance. *'May I?'* is usually a polite request for permission to invade a person's space, e.g. by taking away an unoccupied chair or performing some action that involves touching them. *'You shouldn't have'* shows feigned disapproval to indicate how pleased one is with an unexpected gift (cf.

Hankamer (1978), Pullum (2000), Miller and Pullum (2013). ‘*Können wir []*’ is used in situations where it can be completed with the verb ‘*gehen*’ (*go*) or ‘*anfangen*’ (*start*).¹⁰⁶

Sometimes an ellipsis seems to be exophoric in the immediate context, but is actually cataphoric when the missing verb is presented later in the text after a clarification question in a dialogue. The following example involves the opening lines from a joke (Yvel, 2001: 94). Exophoric, anaphoric and cataphoric reference as well as syntactic ambiguity are typical devices used as a source of misunderstanding in the language of humour:

(5:131) *After creating heaven and earth, God created Adam and Eve. And the first thing he said was: “Don't.” “Don't what?” Adam replied. “Don't eat the forbidden fruit,” God said.*

¹⁰⁶ Some scholars have argued that verbal ellipsis firmly resists exophoric use. Pullum (2000), for instance, described the following scenario: “Suppose you and I come upon a place where just a week ago we admired a beautiful church, and find it is almost completely demolished, and you stare at me incredulously as if I might know why this has been done. [...] In the context just described it would be bizarre for me to say #*Don't look at me; I don't know why they have*. A demolition scene is not sufficient to permit *why they have* to be understood as ‘why they have demolished that beautiful church’, or ‘why they have perpetrated this heinous architectural crime’, or whatever.” In many cases the pronoun ‘*it*’ and / or the verb ‘*do*’ and their German equivalents as proforms are usually grammatically necessary, but some scenarios for true exophoric verbal ellipsis in written and spoken language have been constructed in the literature or cited from creative texts such as advertisement slogans. A very successful advertising campaign for hair colouring products used an elliptical slogan that became a national catch-phrase: ‘*Does she... or doesn't she?*’ This question was answered by the subhead: ‘*Haircolor so natural only her hairdresser knows for sure.*’ This example from a hypothetical scenario of gossip was discussed by Schachter (1977:763) in a slightly shortened form (‘*Does she or doesn't she? Only her hairdresser knows for sure.*’) as being an example of exophoric verbal ellipsis. Its full structure (‘*Does she dye her hair or doesn't she dye her hair?*’) is deducible from the nature of the product being advertised and the photograph of a model with voluminous, healthy-looking hair and a beautiful hair colour.

What some authors would call cataphoric lexical verb ellipsis (5:132-135) is better analysed as a type of Right Node Raising (RNR, cf. Chapter 5.1.8). We find Right Node Raising in some German corpus texts, with the non-finite verb at the end of the sentence, and rarely in the English texts where gapping or other structures are preferred (5:133b). Both RNR and gapping involve the deletion of lexical verbs in coordinated clause complexes. They also share some aspects of focus and prosodic structure (Féry and Hartmann, 2001).

(5:132) *Bei der Entscheidung für die Währungsunion hätte nicht die*

ökonomische Einsicht, sondern die Politik dominiert.

(GO_ESSAY_003)

(5:133) a) *Durch diese Anschläge wird die Stromversorgung für Sie und*

Ihre Familien und werden die Öleinnahmen für alle Iraker

unterbrochen. (GTRANS_SPEECH_004)

b) *These attacks deny electricity to you and your family and oil*

revenues to all Iraqis. (EO_SPEECH_004) (-> gapping)

In a written text without prosodic cues, the structure examples such as (5:134-135) can be locally ambiguous, particularly as the reader may at first look for potential antecedents of the first modal verb in the previous sentence. (5:134-135) show that RNR, like gapping, involve some semantic contrast (*can* vs. *cannot*; *wollten* vs. *konnten*). Additionally, we

have two grammatical subjects that are also contrasted (*Jim* vs. *Jerry*; *sie* vs. *wir*). In spoken language, one of these verbal or nominal pairs of contrasted elements would usually involve a clear prosodic contrast (one element carrying a pitch accent and the other element being deaccented). The first part of the structure cannot be separated from its second part.

(5:134) *Jim can but Jerry cannot make the meeting.*

(5:135) *Sie wollten und wir konnten nichts dagegen tun.*

As explained in the previous chapter, endophoric nominal ellipsis can occur as a clause-internal phenomenon or even within a phrase. Endophoric verbal ellipses-antecedent relations typically are cross-clausal or cross-sentential anaphoric relations. Some structures that Halliday and Hasan call lexical verb ellipses or operator ellipses within the sentence (1976: 174f.) are verbs that we may see as part of the same phrase. They involve sequences of modal, auxiliary or lexical verbs showing contrasts in tense, modality, aspect or negation, associated with only one grammatical subject. It is not necessary to characterise the following examples as ellipses. The modal verb and its negative form are a coordinated sequence that does not necessarily involve an ‘omission’ of the lexical verb (5:136). Similarly, in (5:137) there is no need to call this structure an ‘omission’ of the operator.

(5:136) *They might or might not have objected.* (identified as lexical verb ellipsis by Halliday and Hasan, 1976: 175, although the lexical verb is related to two or more coordinate ‘operators’).

(5:137) *They must have been both watching and being watched* (called ‘operator ellipsis’ within the sentence in Halliday and Hasan, 1976: 174).

Similar structures are illustrated in (5:138-141) with examples for our corpus data, where a lexical verb is related to a sequence of auxiliary or modal verbs. Particularly in English, the second verb in such a sequence is sometimes understood as a parenthetical expression and set off by commas or, more forcefully, by dashes or parentheses. This is a rhetorical figure commonly used in informative and persuasive texts to add some specific information or to emphasise the verb phrase. As mentioned above, some would call this cataphoric verb ellipsis, but we do not consider such structures to be actual omissions.

(5:138) *Einen moralischen Schlussstrich kann und darf es nicht geben.*
(GO_SPEECH_009)

(5:139) *At all times we have and will put stability and the national economic interest first.* (EO_ESSAY_005)

(5:140) *Trade and environmental objectives can, and must, be complementary.* (EO_ESSAY_007)

(5:141) *Power in the service of freedom is to be welcomed, and powers that share a commitment to freedom can -- and must -- make common cause against freedom's enemies.* (EO_SPEECH_012)

A slightly comparable corpus example is (5:142) where the author switches from active to passive voice in the verb phrase and relates both forms of the same lexical verb to the same subject and object serving both as agent and patient. Adding overt linguistic material would be possible, so that the object could appear twice in the full sentence, but it need not be interpreted as an incomplete structure.

(5:142) *The United States affects and is affected by developments around the world.* (EO_ESSAY_001)

As discussed in the previous chapter, all cross-clausal nominal ellipses are included in the analysis of cohesive cases due to the reasons explained above. Endophoric verbal ellipses can occur across clauses in clause complexes, but if they are the direct result of clause coordination or subordination they are not primarily a cohesive phenomenon. Gapping, for instance, is a structural, and not a cohesive relation. Halliday and Hasan exclude all cases of gapping from ellipses as cohesive devices even if they

occur across sentences boundaries (5:143) as such sentences are always linked by a coordinating relation.

(5:143) *The cat won't catch mice in winter. Nor the dog rabbits.* (example from Halliday and Hasan, 1976: 203)

Gapping is predominantly found in coordinate clauses belonging to clause complex (5:144-145).

(5:144) *Max went to the store, and Oscar [] to the arcade.*

(5:145) a) *But since 1992, the British economy has grown some 40%, the German [] barely [] 10%. (EO_ESSAY_012)*

b) *Aber seit 1992 ist die britische Wirtschaft um rund 40 Prozent gewachsen, die deutsche [] nur um knapp zehn Prozent []. (GTRANS_ESSAY_012)*

Given the fact that gapping is an anaphoric relation that occurs across clause boundaries, it is worth exploring this phenomenon in slightly more detail. In my opinion, it is a type of non-cohesive verbal ellipsis, restricted by syntactic principles. Gapping usually applies to only one constituent in a coordinated clause complex where a simple lexical verb or compound

verb phrase is omitted.¹⁰⁷ As in (5:145a&b), gapping can co-occur with a cross-clausal nominal ellipsis in both English and German to keep the second conjunct as short as possible (cf. Chapter 5.3.5 on mixed cases). The omitted verb in gapping prototypically is a finite lexical verb in ‘clause-medial position’ (Repp, 2009b: 26), but the additional omission of auxiliaries or the entire verb phrase is also possible in gapping structures as (5:144-145) demonstrate. Therefore, if a complex verb phrase is omitted in gapping, auxiliaries, modals and main verbs are deleted from different positions, e.g. before and after an adverb in English (5:145a) or in second and final positions in German (5:145b). As noted by Kuno (1976: 306), what is deleted by gapping can be more than just the verb. It may also involve verb-phrase internal material such as the direct object as complements. The second conjunct has to contain two contrastive remnant constituents (e.g. a subject noun phrase and an object or an adverbial adjunct) and a ‘gap’, which is interpreted as identical to the verb(s) in the first conjunct. The following gapping example (5:146) which has been taken from Kuno (*ibid.*) has different readings without context. In spoken language, different prosodic patterns are possible which will lead to

¹⁰⁷ The underlying grammatical structure of gapping seems similar to what can be found in many languages, even in Ancient Greek (e.g. Luke 17:31: ἐν ἐκείνῃ τῇ ἡμέρᾳ ὃς ἔσται ἐπὶ τοῦ δώματος καὶ τὰ σκεῦη αὐτοῦ [] ἐν τῇ οἰκίᾳ, μὴ καταβάτω ἄραι αὐτά. – In that day, he which shall be upon the housetop, and his stuff [] in the house, let him not come down to take it away. – English-Greek Parallel New Testament, p. 268 https://books.google.de/books?id=DDP14C8_TwAC&pg=PA268, cf. <http://biblehub.com/text/luke/17-31.htm> for a word-for-word transliteration [last checked 08/02/2016] – Possibly, the earliest cases of gapping are constructions like this example where the copula verb was omitted in the second conjunct. According to Krisch (2009: 191) gapping is a type of ellipsis that can be found not only in Latin and Greek, but also in Sanskrit and Hittite, and can even be reconstructed for Proto-Indo-European.

different interpretations (Winkler, 2006: 421, see also Winkler's habilitation thesis from 2003).

(5:146) *My brother visited Japan in 1960, and my sister [visited Japan] in 1961.*

All forms of gapping are characterised by the omission of the posterior member of a pair of lemma-identical verbs. Gapped sentences resemble answers to implicit multiple wh-question (e.g. *Who reads what?*):

(5:147) *I tried to read Aspects, and John [] LGB.* (example from Vanden Wyngaerd, 1998:33)

The two conjuncts in gapping are related by the non-syntactic principle of contrast, but they have to be syntactically as symmetric as possible. Otherwise such constructions often do not make an unambiguous interpretation of the elliptical conjunct possible or are ungrammatical (cf. Chapter 3.3 in which possible ellipsis-antecedent mismatches in gapping structures are discussed). Cases such as the following sentence where nominal phrases are contrasted with prepositional phrases or adjective phrases are not possible or are at least very rare in gapping structures and only seem to occur when predicative expressions are contrasted and a form of 'be' is omitted (5:148-149).

(5:148) **John eats apples and Peter [eats] in the car.*

(5:149) *The story was demanding and she [] its single parent.*¹⁰⁸

For English, pseudogapping structures, which share some important similarities with gapping, have been discussed extensively in the literature (e.g. Levin, 1978, Lasnik, 1999b, Hoeksema, 2006, Merchant, 2008, Gengel, 2013). Pseudogapping is viewed as marginal, informal and as a point of variation among speakers of English. In some textbooks it has been imprecisely defined as “a variety of ellipsis where the accusative object is not deleted” (Carnie, 2013: 236), but it can be better described as an omission that characteristically has a finite auxiliary in front of the ellipsis site (or a form of the substitutional dummy-verb ‘do’) and some contrastive material (typically a noun phrase or a prepositional phrase) following the ellipsis site (Gengel, 2013: 9). As does gapping, pseudogapping obeys locality restrictions (5:150-153).

(5:150) *He realized he could make more money in some other position than he can farming.* (example from Levin, 1978)

(5:151) *John could pull you out of a plane, like he did his brother.*
(from Sag, 1976)

¹⁰⁸ Example taken from the novel *Purity* (2015) by Jonathan Franzen, p. 222, <https://books.google.de/books?id=cT94BwAAQBAJ&pg=PA222> [last checked 04/02/2016]

(5:152) *That may not bother you, but it does me.* (from Hoeksema, 2006)

(5:153) *Although John wouldn't give Bill the book, he would Susan.* (from Takahashi, 2004)

Such structures are rare in our corpus data. (5:154-155) seem to be the only examples of pseudogapping in the entire corpus.

(5:154) *It was a singular object that somebody treasured almost the way they would an oil painting.* (EO_ACADEMIC_004)

(5:155) *Start your online process for each package as you would with any shipment.* (EO_WEB_009)

Pseudogapping in English involves a typical syntactic pattern with the finite auxiliary in a medial position between the subject and a third constituent. In German, a 'pseudo-gap' would not necessarily occur in the same position as the position of the finite verb depends on the clause type. Standard German probably does not have pseudogapping at all, although one example that seems rather similar to pseudogapping structures has been suggested by Winkler (2013: 481, citing a passage from Anna Seghers' book *Das siebte Kreuz* [5:156]).

(5:156) *Aber ich kann sie mir nicht mehr vorstellen, dachte er. Wallau
kann ich und alle andern.*

This example is still acceptable, but it is grammatically marginal in German and probably has been written in this way to represent a spontaneous non-standard pattern of spoken language. The construction also seems semantically odd in this context, as the modal verb 'können' in combination with an accusative object but without a lexical verb is often used as a main verb (as in 'Er kann das Gedicht.' / 'Er kann gar nichts.' / 'Er kann Deutsch. -> It is possible to add a verb or to analyse 'können' as a synonym of 'fähig / in der Lage sein, etwas auszuführen; etwas beherrschen'),¹⁰⁹ which normally blocks 'können' + lexical verb ellipsis if an accusative object follows (cf. Chapter 5.2.4). English pseudogapping actually occurs across clauses in clause complexes, which is not possible in German: '*Obwohl ich sie mir nicht mehr vorstellen kann, kann ich (mir) Wallau [].' Verb valency, reflexivity of the main verb and its more frequent use with an abstract or inanimate object instead of a person ('sich etwas vorstellen') can involve a complex set of additional reasons for the ungrammaticality of this sentence.¹¹⁰

¹⁰⁹ cf. <http://www.duden.de/rechtschreibung/koennen>, 2. [last checked 04/02/2016]

¹¹⁰ The English translation Winkler gives for the above cited German construction which involves lexical verb ellipsis, but not pseudogapping, is: 'But I can no longer picture her. I can Wallau, and everyone else.' The official English translation by James A. Galston (*1), which apparently appeared in 1942 before the German original (*2) was: 'I can't picture her to myself any more, he thought. Wallau, yes, and all the others too.' Here, the English syntactic structure actually is the one that German would normally use in this case with 'ja' or another particle such as 'durchaus' or 'schon'. In Halliday and Hasan's classification, this would probably fall under clausal substitution (in analogy with 'I think

Bearing all these facts in mind, we come to the conclusion that gapping as well as similar structures such as stripping and pseudogapping are primarily structural and not cohesive relations. They will fall under non-cohesive verbal ellipsis in the analysis. This also includes cases with no explicit coordination or subordination marker (5:157) or with a punctuation mark that splits a coordinated structure. Although occurring across clauses, such omissions are subject to locality restrictions¹¹¹ and depend on the syntactic *relation* between two clauses within a clause complex. They do not have the potential to connect larger passages of text.

(5:157) *The coast, which includes the mysterious and unique Chesil*

Beach, is spectacular - the Landscape [] diverse.

(EO_TOU_006)

so’ or *‘I think (,) yes’* although after the latter, which is less frequent in English (cf. BNC and COCA: <http://corpus.byu.edu/bnc/> & <http://corpus.byu.edu/coca/> [last checked 04/02/2016])¹¹⁰, and the example from Galston’s translation more material can be added, which is not possible in the case of *‘so’*.) Both structures *‘I think so’* / *‘I think, yes’* function as an adverb of certainty or a pragmatic marker which qualifies the truth and probability of the previous utterance (cf. Chapter 6.3 on answering particles).

*1 In: *Early 20th Century German Fiction: A. Döblin, L. Feuchtwanger, A. Seghers, A. Zweig* (2003) by Alexander Stephan (ed.), p. 187, <https://books.google.de/books?id=MS3fBxVg-Z4C&pg=PA187> [last checked 04/02/2016].

*2 cf. G. P. Butler (1998). ‘Pro Captu Interpretis ...: James A. Galston and the Road to Fame’. In: Ian Wallace (ed.). *Anna Seghers in Perspective*. Amsterdam / Atlanta: Rodopi, 1998, p. 93.

¹¹¹ Murguía constructed a few examples that do not display strict locality effects with some material intervening between the elided element and the antecedent in pseudogapping and gapping. She claimed that the relation between the antecedent and the ellipsis in those structures is ‘non-local’ if the antecedent and the elided constituent can be separated by intervening material so that the gap is embedded (Murguía, 2004: 158; 2005:175) as in *‘Mary accepted the job offer, and I believe Peter did too.’* or *‘Tom talked to his wife, and I heard Beth did to her husband.’*

Similarly, other cases of verbal ellipsis in pairs of clauses that are connected by other coordinating or subordinating relations all fall under the umbrella of being a ‘non-cohesive’ ellipsis if the verbal ellipsis heavily depends on these structures. We annotated these cases to explicitly distinguish them from cohesive verbal ellipses and we have listed them under the heading of ‘non-cohesive ellipsis’. It may, at first, not seem a straightforward matter to decide which examples of verbal ellipsis actually increase textual cohesion and which should count as non-cohesive. The exact impact of verbal ellipsis on textual cohesion is rather difficult to determine and we recognise that it is a debatable issue. While Halliday and Hasan only looked at textual ties which stretch beyond the sentence, this would be a rather narrow definition of cohesion in general as Hoffmann confirms (2012: 73). Hoffmann, for instance, excluded all clause-internal relations in the same way we did, but included all relations between clauses and sentences. Cross-clausal relations of some cohesive devices such as lexical cohesion or nominal ellipsis do not heavily depend on coordinated or subordinated clause structures and therefore, in my opinion, contribute to the cohesiveness of a text. They are relatively unaffected by the exigencies of sentence structure and can establish inter-sentential relationships and true text-forming relations. In the case of cross-clausal verbal ellipsis, this is different. Coordinating or subordinating conjunctions can license verbs to be elided from a clause if they have a counterpart in a previous clause (5:158-159).

(5:158) *We never get it all “right” in any year, and probably never will [].* (EO_SHARE_004)

(5:159) *Better go on up while you still can [].* (EO_FICTION_006)

All endophoric ellipsis-antecedent relations that go beyond clause or sentence boundaries have been annotated as verbal ellipsis in a first annotation step to achieve consistency and comparability for all annotated cohesive devices in the GECCo project. In a second step, those examples that are better analysed as the result of coordination or subordination and cannot refer back to textual passages longer than the coordinated or subordinated structure have been filtered out and listed separately. Non-cohesive ellipses will be analysed for comparative purposes and to clearly distinguish them from cohesive ellipses.

Verbal ellipses that contribute to the cohesiveness of a text fall into two categories in Halliday and Hasan (1976), ‘operator ellipsis’ and ‘lexical ellipsis’, which will be explained in the next sections.

5.2.3 Operator ellipses

This section and the next investigate the concepts of operator ellipsis and lexical verb ellipsis that have been suggested by Halliday and Hasan (1976) as subtypes of verbal ellipsis which can function as cohesive devices. Verbal ellipses, especially those that can function as a cohesive device, prototypically involve the omission of a lexical verb as the head of the phrase. Under certain circumstances, operators can be omitted while the lexical verb remains as a remnant structure ('operator ellipsis'). As Halliday and Hasan's textbook on cohesion in English attempted to cover all types of cohesive devices, the characteristics of lexical verb ellipsis and operator ellipsis are covered in a rather sketchy overview of about five pages (1976: 170ff.) making virtually no reference to earlier literature or the literature of that period on verb ellipsis.

In operator ellipsis, a modal or auxiliary verb is omitted from a verb phrase. It has been called 'ellipsis from the left' (Halliday and Hasan, 1976: 174). The operator is a finite form of '*be*' or '*have*', a modal verb or '*do*' (cf. also Downing, 2014: 21). We see operator ellipsis as an omission of the operator verb, but Halliday and Hasan have also described it as the omission of what they call the 'modal block', i.e. the subject and the finite verbal operator (ibid.: 193).

A few cases of operator ellipsis in coordinated structures fall under gapping, particularly in German with the verb '*werden*' as a passive or future auxiliary:

(5:160) *Some were laughing and others [] crying.* (example from Halliday and Hasan, 1976: 174)¹¹²

(5:161) *Es wurden wirklich viele Rekorde aufgestellt und einige [] nur sehr knapp verpasst.* (GO_FORUM_001)¹¹³

(5:162) *Und Katja wird nicken, und Aleksej [] auf seine Schuhe schauen.* (GO_FICTION_002)

Operator ellipsis in question-answer sequences where the reply looks like a non-finite verbal group as in (5:163-64) is supposed to be very frequent in English according to Halliday and Hasan (ibid.: 191).

(5:163) *What should she have done? – [] Told the police.*

(5:164) *Has he been crying? – No, [] laughing.*

In German, omissions of the subject and a modal verb or auxiliary are possible in question-answer pairs or other adjacency pairs in conversational turn-taking and the remnant can be an infinitive or a past

¹¹² Not all of the examples of ellipses that are given in this chapter on verbal ellipses are cohesive cases. I also include some non-cohesive cases for a general illustration of the different types of verbal ellipses. These types can be used as cohesive devices, but they are not always used in this function. This depends on the surrounding syntactical structures (cf. Section 5.2.2 on the distinction between cohesive and non-cohesive cases).

¹¹³ also includes a nominal ellipsis

participle (5:165-166), sometimes accompanied by complements, while in English, apart from these options, it can also be a present participle if the full phrase involves a progressive form.

(5:165) *Hat er geweint? – Nein, [] gelacht.*

(5:166) *Was muss ich jetzt machen? – [] Deine höchste Karte spielen.*

In longer verb clusters, not only the first operator, but all the subsequent operators may be omitted (5:167). The lexical verb, however, must stay intact. In German, equivalent structures often have to be formulated in a slightly different way due to a different structure of the verb phrase.

(5:167) *What could he have been doing?*

[] *Been going to swim, I think.*

[] *Going to swim, I think.*

[] *Swim, I think.* (example from Marcus, 1982: 47)

Halliday and Hasan included operator ellipsis into their taxonomy of inter-sentential cohesion, but there are almost no occurrences of cross-sentential operator ellipsis in our corpus data, which may partly be due to the fact that only some corpus registers – and not all texts within them – are characterised by much turn-taking and numerous adjacency pairs. If cross-sentential operator ellipses as cohesive devices also involve the

omission of the subject, as in the examples given above, they are better analysed as clausal ellipsis (cf. Chapter 8). We only annotated ellipses ‘within the verbal group’ (ibid.: 167) as verbal ellipsis.

Halliday and Hasan (1976: 174) claimed that operator omissions within the sentence are very frequent. As explained above, what Halliday and Hasan call ‘ellipsis’ of the operator in the context of coordination (as in example [5:137] above: *They must have been both watching and being watched*) – will not primarily concern us here if operators are structurally related to complex verb phrases. This does not have to be seen as an omission. Non-repetition of subjects and auxiliary verbs can be considered the norm when clauses with the same subject are coordinated (cf. example [8:1] below in Chapter 8: *He was neat and tidy and determined to get forward at every chance*). The insertion of the subject and the auxiliary verb in every position where it is theoretically possible within a clause complex would be redundant and not the grammatical norm or the grammatically unmarked structure.

Sometimes an optional additional element is added to an essentially complete structure, which will then again make it seem like an omission. In the following German example (5:168), ‘*es*’ has been inserted within the verb phrase to make the second part of the verb phrase look like a separate clause where ‘*werden*’ should be repeated in a complete structure.

(5:168) *NIEMAND in Dortmund hat je geleugnet, dass man nicht Meister werden kann oder es [] will. (GO_FORUM_001)*
(vs. *NIEMAND in Dortmund hat je geleugnet, dass man nicht Meister werden kann oder will.*)

English comparative constructions also fall under structures where optional constituents may be added and trigger verbal ellipsis, if we assume a fully-fledged underlying syntactic structure which is parallel to the antecedent clause. They fall under lexical verb ellipsis (see Chapter 5.2.4).

Verbal ellipses that function as cohesive devices are prototypically linked to content words as textual antecedents. Operator ellipsis as an ellipsis within the verb phrase is rare in our data. Text-type-specific or informal auxiliary drop (cf. Chapter 6.5) or the omission of function words in general as a reduction strategy usually do not contribute particularly to increased cohesiveness of a text.

5.2.4 Lexical verb ellipses

After modal verbs and after the auxiliaries ‘*be/sein*’, ‘*have/haben*’, and ‘*werden/will*’, lexical verbs can be left out in both English and German. In the annotation process, structures are annotated as lexical verbal ellipsis remnants if the lexical verb is missing in the verb phrase. Additionally, lexical verb ellipsis in the annotation scheme can include the omission of verb phrase complements (e.g. objects or adverbial phrases). In other words, those cases where additional material that belongs to the *extended verb phrase* (cf. p. 177 above and Huddleston, 1984: 112) is omitted in addition to the lexical verb fall under verbal ellipsis in the analysis. However, if the lexical verb *and* the subject have been omitted together, this will rather fall under clausal ellipsis in the annotation as the subject does not belong to the extended verb phrase (5:169). In (5:169), both the German and the English sentence are ambiguous without further context as ‘*Dorcas*’ can be interpreted to be the subject or the object of an elliptic clause.

(5:169) a) *I hate that stuff – Dorcas [] too.* (EO_FICTION_001)

b) *Ich kann so was nicht leiden – und Dorcas [] auch nicht [].*

(GTRANS_FICTION_001)

Omission of optional valency complements or other optional elements that only involve semantic implications will not be annotated if no verbal element has been deleted from the verb phrase.

Lexical verb ellipsis can be accompanied by ellipsis of one or more than one operator (5:170-71), with the exception of the first element of the verb phrase. It has been called ‘ellipsis from the right’ as the omission of the lexical verb ‘may extend “leftward” to leave only the first word intact’ (Halliday and Hasan, 1976: 173). It is rare in our data that lexical verb ellipsis refers to long verb phrases involving several non-lexical verbs which is partly due to the fact that question-answer turns are not extremely frequent in the corpus data.

(5:170) *Could she have been going to swim?*

Yes, she could have been going to swim.

Yes, she could have been going to [].

Yes, she could have been [].

Yes, she could have [].

Yes, she could [].

(5:171) *Has anyone been eating? – Jane has [].*

The underlying structure of anaphoric lexical verb ellipsis does not have to mirror the complete structure from a previous clause. In (5:172), in contrast to (5:170), the modal verb is not part of the verb phrase in the reply.

(5:172) *A: You should have been waiting here when the taxi arrived. –*

B: I have been [] all the time since lunch. (example from Quirk et

al., 1985: 52)

Lexical verb ellipsis has language-specific features. In English lexical verb ellipsis, we often find ‘stranded’ auxiliary verbs, modal verbs at the end of a clause (5:173-176). English can also have lexical verb ellipsis in sentences ending with the infinitival complementiser ‘to’ (5:176), but German has no sentences ending with infinitival complementisers.

(5:173) *The first dispute sparked by the competition between bells and muezzin: Has the Islamicization of the Orient reached, or passed, its zenith? Abbas thinks it has [].* (ETRANS_FICTION_005)

(5:174) *Will the United 26 finally forget or abandon the UK? because of its irrelevance. I think everyone in the UK is desperately hoping they will [].* (EO_FORUM_003)

(5:175) *It just doesn't switch'em? Yeah it doesn't [].*
(EO_ACADEMIC_006)

(5:176) *They came to visit one Christmas time and we realized how serious the problem was when Dad couldn't get out the chair. And in the morning he'd been able to [], come the afternoon he couldn't [].* (EO_INTERVIEW_003)

In German translations, there is a stylistic preference or the syntactic obligation to insert 'es / dies / das' + 'tun' (5:177b, 5:178b).

(5:177) a) *We should at least make the effort to try. The political and economic stakes involved in this case are simply too great not to [].* (EO_SPEECH_009)

b) *Wir sollten es zumindest versuchen. Politisch und wirtschaftlich steht einfach zu viel auf dem Spiel, um das nicht zu tun.* (GTRANS_SPEECH_009)

(5:178) a) *We can no longer accept the level of failure that we have [] in the past and this legislation says that we won't [].*
(EO_SPEECH_010)

b) *Wir können Misserfolge auf diesem Niveau nicht weiterhin akzeptieren, und diese Gesetzesvorlage bedeutet, dass wir das nicht mehr tun werden.* (GTRANS_SPEECH_010)

German sometimes uses ellipsis of the whole verbal group – which rather falls under the category of clausal ellipsis as an even more reduced structure – to avoid lexical verb ellipsis within a verbal group (e.g. [5:179] 'Aber ich.' -> 'But I can.'). Nevertheless, lexical verb ellipsis can be used in certain similar contexts in English in German, particularly after modal verbs in spoken language or fictional dialogues (5:179).

(5:179) a) *Kannst du übrigens über das Haus springen? Nein? Aber*

ich []. Über das Haus und über einen Baum. Soll ich []? Mach doch! Ich könnte ja leicht [], aber ich will nicht []. So, du willst nicht []? (GO_FICTION_010)

b) Say, can you jump over the house? No? But I can []. Over the house and over a tree. Should I []? – Go ahead! – I could real easy [], but I don't want to []. – Oh, you don't want to []? (ETRANS_FICTION_010)

German sometimes appears to have a subject-auxiliary inversion in sentences that only consist of the subject and an auxiliary or modal verb and involve topic drop (5:180-81).¹¹⁴ These are problematic cases with regard to the ellipsis category they fall into; they can be interpreted in two ways.

(5:180) a) Hast du etwas gegessen? – Ja, [das] habe ich [getan / gemacht]. / ? Ja, [etwas gegessen] habe ich.

b) Have you eaten something? – Yes, I have [eaten something / done this].

(5:181) Die werden mit müden, heiseren Stimmen gegen das Urteil protestieren, [das] sollen sie [tun]. /... [?protestieren] sollen sie. (GO_FICTION_001)

¹¹⁴ Structures similar to English lexical verb ellipsis after auxiliary verb can occasionally be found in colloquial German as in the following (ungrammatical?) extract from an online conversation I recently came across: A: *Ohne Motorhaube rumgefahren* – B: *Nur kurz.* – A: *Aber wir sind []!*

b) *They'll protest the verdict with hoarse, weary voices, let them*
[protest / do this]. (ETRANS_FICTION_001)

In an interpretation as anaphoric verbal ellipsis, the ellipsis site can be filled with a verb from a previous sentence. On the other hand, it can be completed with proforms that substitute the antecedent. The first position in German can be completed with *'dies / das'*. A form of *'tun / machen'* can be added after the subject. English can also repeat the verb phrase from the previous sentence or add a form of *'do'* and a pronoun. This would be a highly conventionalised ellipsis where semantically relatively empty or unspecific elements fill the ellipsis site regardless of the respective context. It is not considered necessary to describe such constructions here in great detail as they are a marginal phenomenon in our corpus data. They more often occur in colloquial, everyday speech with many turn-takings.

In English and German, lexical verb ellipsis sometimes co-occurs with *'too/as well/also/either'* or negative forms as explicit markers of parallelism or contrast, cf. examples (5:128-129) above and (5:182). If modal verbs or auxiliaries are left as remnants, proform are usually added in German.

(5:182) *Nintendo works hard to protect your privacy – and you should [],*
too. (EO_WEB_001)

Thomas (1987) described two subtypes of lexical verb ellipsis: 'verb

group echoing' and 'auxiliary contrasting ellipsis' which either involve echoing or contrasting auxiliaries or modals. A different auxiliary or modal verb can be chosen for contrast with the antecedent or tense or polarity of the same auxiliary can be new in the contrasting sequence followed by lexical verb ellipsis. A contrasting auxiliary will typically be stressed in spoken language. Polar questions begin with an auxiliary or modal verb and can be answered using the same verb or a different one. In English, we typically find lexical verb ellipsis in such contexts (5:183-5:184).

(5:183) *Has he visited India? – No, he hasn't.*

– No, but he will / might [] soon.

(5:184) *Has the Islamicization of the Orient reached, or passed, its zenith? Abbas thinks it has [].* (ETRANS_FICTION_005)

We find a few examples of lexical verb omissions after auxiliaries and modals in our corpus data that do not refer to a textual antecedent and can rather be seen as an interruption or breaking off in mid-sentence (5:185-86), cf. Chapter 6.6.

(5:185) a) *Who told you? I didn't ... I never ...* (EO_FICTION_009)

b) *Wer hat dir das gesagt? Ich habe nicht ... Ich bin nie ...*

(GTrans_FICTION_009)

(5:186) a) *Then, deliberately, he turned back to the woman on the stool*

next to his. 'I don't have a car,' he said. 'If I borrowed one, would you... ?' (EO_FICTION_009)

b) *Dann wandte er sich demonstrativ wieder der Frau zu, die auf dem Barhocker neben ihm saß. „Ich habe kein Auto“, sagte er.*

“Wenn ich mir eines leihen würde, würdest du ...?“

(GTRANS_FICTION_009)

A typical case of lexical verb ellipsis is ellipsis in comparative constructions. It is sometimes possible to add an additional verb or more material to a comparative construction to form a clausal structure although the sentence would not necessarily be incomplete without it. ‘*Than*’ and ‘*as*’ and their German equivalents ‘*als*’ and ‘*wie*’ can be used both as conjunctions or prepositions and can combine with clausal or with phrasal complements, which has led to a distinction between clausal comparatives and phrasal comparatives (Hankamer, 1973). German with its relatively free word order mainly uses phrasal comparatives that can be put in front of the lexical verb where the same information in English is put in a clause involving an ellipsis of the lexical verb (5:187).

(5:187) a) *Und wenn die EU sich wie in der Vergangenheit verhält und*

dem Airbus Finanzierung zu Zinssätzen unter den auf dem Markt

gültigen bietet, könnte uns ein großer und sehr kontroverser

Kampf in der WTO bevorstehen. (GTRANS_SPEECH_009)

b) *And if the EU does as it has [] in the past, and provides financing to Airbus at below-market rates of return, we could be facing a very large and highly contentious fight in the WTO.*
(EO_SPEECH_009)

Particularly in English texts, we find optional repetitions of modal verbs or forms of 'be' / 'have' after comparisons that have been inserted to create a separate sub-clause (5:188).

(5:188) a) *He can run faster than Jane can [].* vs.

b) *?Er kann schneller rennen, als Jane [] kann.*

Such structures with a repetition of a modal or auxiliary verb are interpreted as clausal comparatives with lexical verb ellipsis. When a form of 'do' is inserted in a comparative construction instead of another modal or auxiliary verb, its analysis is less straightforward. In the English literature, this use of 'do' has been described as a reduced clause 'where DO acts as a dummy operator preceding ellipsis of a predication' (Quirk et al., 1985: 134), but at the same time 'do' is seen as a replacement of the verb phrase (5:189).

(5:189) *Mary reads books faster than I do.* [do = 'read books'] (example from Quirk et al., 1985: 134)

It can be difficult for annotators to decide whether this is verbal ellipsis or verbal substitution in the English corpus data (5:190-191, cf. also Chapter 5.2.5). In German, we usually find phrasal comparatives as translations for such constructions (5:191b, 5:192b).

(5:190) *Delta FosB appears to function very differently in addiction than CREB does.*¹¹⁵ (EO_POPSCI_003)

(5:191) a) *Climate change got more ink in a day than nuclear arsenals did in a year.*

b) *Über den Klimawandel wurde an einem Tag mehr geschrieben als über die Atomarsenale in einem Jahr.*¹¹⁶

(5:192) a) *Some amphibians have more than five times as much DNA as mammals do.* (EO_POPSCI_004)

b) *Die Zellen einiger Amphibien besitzen mindestens fünfmal mehr DNA als die Zellen von Säugetieren.*
(GTRANS_POPSCI_004)

For phrasal comparatives, we generally favour an analysis of a non-clausal structure over a deletion or ellipsis analysis in a covert clausal structure. Nevertheless, in English grammar books comparisons such as

¹¹⁵ There is no translation of this sentence in the German corpus data as translations of popular scientific texts sometimes tend to be relatively free.

¹¹⁶ Example taken from the novel *Purity* (2015) by Jonathan Franzen, p. 216 and its German translation *Unschuld* (2015), p. 223, <https://books.google.de/books?id=cT94BwAAQBAJ&pg=PA216&lpg=PA216> / <https://books.google.de/books?id=mirzCAAAQBAJ&pg=PT223> [last checked 04/02/2016]

without repetition of the verb and without 'do'-insertion in a second clause are often called 'ellipsis in shortened comparisons' or 'ellipsis of the predicate' (Huddleston, 1988: 98; Cowan, 2008: 287). According to Lechner who provides a very detailed analysis of ellipsis in comparatives, a sentence such as (5:193) counts as an ellipsis, which in this case can have a wide or a narrow interpretations (Lechner, 2004: 191).¹¹⁷

(5:193) *John wanted to write more plays than Sam* [wrote / wanted to write].

In English, adding a verb is possible if the sentence does not end in an emphatic personal pronoun in its objective case form (e.g. 'me/her/him' 5:194a). After possessive pronouns, it would be possible to expand the structure in order to obtain a complete subordinate clause (5:195). In German, it is less plausible to assume an underlying clause involving an ellipsis of the lexical verb in a second clause as this would have to be indicated by a comma.

(5:194) a) *She runs faster than I* [?]¹¹⁸ / *She runs faster than me* [/]. vs.
b) *Sie rennt schneller als ich* [/].

¹¹⁷ Moore (2001, reference not found but cited as in Carlson, 2002: 122) conducted an interesting corpus analysis and a sentence-completion study to examine how children process comparatives. She found that children most frequently produce sentences with subjects following 'than', but given a specific sentence such as 'Goofy ate more carrots than ___', children more often chose an object ('peas') instead of a subject ('Mickey'). She suggested that children and adults may have different preferred interpretations and different structures for comparatives.

¹¹⁸ '[?]' indicates that it is debatable whether there is an underlying structure in the assumed ellipsis site. '['/' indicates that there is no underlying syntactic structure.

(5:195) a) *Your guess is as good as mine [?].*

b) *Dein Tipp ist so gut wie meiner [/]*

In sentences that involve a comparison and co-referential subjects, it is not possible to reduce the comparative clause to a single element to obtain a phrasal comparative as the verb phrase, and not the subject, involves contrast. The lexical verb in the subordinate clause can be omitted after (or in German in front of) the modal verb (5:196).

(5:196) a) *On an island where people live longer than they should [], the ambitious young Mayor has a plan.¹¹⁹*

b) *Auf einer Insel, auf der die Menschen länger leben, als sie [] sollten, hat der junge Bürgermeister einen Plan.*

Annotators may sometimes face difficulties in identifying lexical verb omissions after modal verbs, as the status of some verbs between modal and lexical verb – and therefore the status of a verb phrase as being incomplete or not – is not always obvious at first sight. The status of modals in the grammar of English and German has never been definitely established as they are subject to ongoing grammaticalisation processes and recent diachronic change. English has four paired modals (*'can' / 'could', 'may' / 'might', 'shall' / 'should', 'will' / 'would'*). Together with

¹¹⁹ example from an advertisement text about a theatre play (*The Island of Immortality*), <http://centralstudio.co.uk/theatre.html> [last checked 28/12/2015]

'must' they are usually referred to as 'central' or 'core' modals (e.g. Quirk et al., 1985: 137), or modal auxiliaries. As members of the larger class of auxiliaries which also includes the 'primary auxiliaries' *'be'*, *'have'*, and *'do'*, the English modal auxiliaries exhibit a number of inflectional and syntactic properties that distinguish them from lexical verbs (Collins, 2009: 12). The so-called NICE properties serve to distinguish primary auxiliaries and modal auxiliaries from lexical verbs. The NICE properties are: negation, inversion, emphasis and 'code' (cf. Huddleston, 1976: 333). Code means that they permit anaphoric ellipsis of the lexical verb. The English core modals are defective verbs and have no *-s* form of the 3rd person singular and no non-finite forms. They usually cannot co-occur in sequences of several modals. Other semantically similar forms or periphrastic equivalents to modal auxiliaries (*'dare'*, *'need'*, *'have to'*, *'be able to'*, *'ought to'*, *'used to'*, *'be allowed to'*, *'be supposed to'*, *'want to'*) have been referred to as 'marginal' modals or 'quasi-modals' (Lakoff, 1972: 239, Westney, 1995: 2ff.). Within the group of quasi-modals, we also find modal idioms or 'semi-modals' (Collins, 2008: 16). Only a few marginal modals and quasi-modals permit anaphoric lexical verb ellipsis (mainly after infinitival complementiser *'to'*). They can be placed on a grammatical continuum between modal and lexical verbs.

It is well known that in older stages of the language, English core modal verbs were independent verbs (e.g. in Old English: *'cunnan'*: to know how to, to be acquainted with; *'magan'* ('to be able to', 'have the power to'; *'sculan'* 'to be obliged to', 'have to'). It has been suggested that English is

still undergoing a shift with regard to the usage and features of modal verbs. Bolinger claimed in 1980 that the system of modal auxiliaries in English at that time was in the process of undergoing a ‘wholesale reorganization’ (Bolinger, 1980: 6). Several lexical verbs as emerging modals and periphrastic constructions as quasi-modals have been and still are assuming some features that are typical of the core modal paradigm, which is changing their categorial status. Various recent corpus studies of morpho-syntactic changes in the English verb phrase have suggested a declining frequency of the core modals, particularly in spoken English, while other forms expressing modality are undergoing grammaticalization processes such as syntactic simplification or semantic bleaching and seem to become more frequent (Krug, 1998, 2000; Leech, 2003, Mair and Leech, 2006, Van der Auwera et al., 2013, Leech, 2013). Reduced pronunciation and phonological weakening is another indicative of change in progress and ongoing auxiliarization (e.g. ‘gonna’, ‘gotta’, ‘hafta’, ‘wanna’).

These findings have inspired recent research to examine whether similar grammaticalization processes are going on in German (e.g. Jäger, submitted). It has been claimed that German modals have retained their polyfunctionality and array of different meanings to a stronger extent than English ones. They share more morpho-syntactic similarities with main verbs and, although English and German modal verbs derive from a common source, German modals present a more homogenous verbal class than English ones (Abraham, 2001: 19ff.). While in English modal verbs

have traditionally been defined based on their syntactic behaviour and the applicability of the NICE properties, German modal verbs do not display the same syntactic properties and it is questionable whether they form a syntactic class at all (Reis, 2001) or whether they are in fact auxiliary or main verbs (cf. Diewald, 1999: 50ff.). The German ‘core’ modals are ‘*dürfen*’, ‘*können*’, ‘*mögen*’ (‘*möchten*’), ‘*müssen*’, ‘*sollen*’, and ‘*wollen*’. Some authors have also suggested ‘*brauchen*’, or ‘*werden*’ and some periphrastic constructions as ‘semi-modals’ or ‘modalisierende Verben’ (cf. the discussions in Diewald, 1993: 218, 1999: 50, Reis, 2001: 287, Klein, 2009). Like English modals, German modal verbs are usually not used in the imperative mood or passive voice. As German modal verbs normally take infinitive complements, they can be followed by anaphoric ellipsis of the lexical verb if the infinitive is left out. German modal verbs are polyfunctional and can also often function as main verbs with a specific meaning. Their syntactic behaviour does not distinguish them from main verbs; therefore it can be difficult for annotators to decide whether a German verb is used as a modal verb followed by a lexical verb ellipsis or as a full verb in a grammatically complete verb phrase.

If we assumed that modal verbs behave like auxiliaries and cannot occur without a main verb, we would have to analyse all sentences in which modal verbs are used as the only verb as ellipses, e.g. (5:197-98) (cf. Diewald, 1999: 53).

(5:197) *Kaum eine andere Volkswirtschaft der Welt könnte das [?]. Die deutsche kann es [?].* (GO_SPEECH_012)

(5:198) *Ich trat auf eine Platte, irgendwas von Udo Jürgens. „Oh, Mist! Das wollte ich nicht [?].“* (GO_FICTION_006)

German modals actually can sometimes occur in the position of the lexical verb. Some constructions originally involving lexical verb ellipses after modal verbs have become rather conventionalised phraseological units in examples such as *‘Ich muss zum Arzt’* and seem to be analysed in a similar manner as certain forms of implicature or to be grammatically reanalysed as a main verb (cf. Duden entry where this example is listed under *‘müssen’* as a main verb in the sense of ‘to be obliged to go somewhere’). Zifonun et al. (1997: 1256) confirm that this construction should not be explained with an ellipsis of a verb of movement as it has undergone grammaticalization. For other German modal verbs, the Duden also lists certain usages as main verbs,¹²⁰ which some native speakers may still consider to be elliptical modal verbs or at least formulaic structures involving a conventionalised ellipsis interpretation, cf. also the discussion of example (5:156) above. They do indeed resemble lexicalised ellipses with a standard interpretation regardless of the textual context, but may have undergone or still undergo a process of ‘degrammaticalization’ to

¹²⁰ <http://www.duden.de/rechtschreibung/duerfen> / <http://www.duden.de/rechtschreibung/koennen/> http://www.duden.de/rechtschreibung/wollen_moechten_wuenschen [last checked 19/02/2016]

acquire specific lexical meanings while newer lexical and grammatical meanings of these forms co-exist ('layering', see Hopper and Traugott, 1993: 124). This may be a counterexample to the often assumed unidirectionality of grammaticalization and an indicator of cyclical language (cf. Abraham, 2010, on cycles of grammaticalization). Usually, reanalysis in grammaticalization is claimed to be unidirectional so that lexical verbs become auxiliaries but not the other way round. Another example against the unidirectionality hypothesis is the development of a modal auxiliary to a lexical verb in Pennsylvania German¹²¹ ('*wotte*' – in the sense of 'to wish' as a lexical verb derived from the modal auxiliary '*welle*' ['want to'], cf. Burridge, 1998).

Zifonun et al. (1997: 1255) mention a continuum on which German modal verbs can be placed ('fließende[r] Übergang zum Vollverbcharakter'). If the function of a single modal verb is not entirely clear in the corpus data where it can be seen as a main verb, a formulaic sequence involving conventionalised ellipsis or an anaphoric elliptical verb phrase as in example (5:199a), it can be marked as 'problematic' in the annotation. In the English original (5:199b), this example more clearly involves lexical verb ellipsis after the infinitival complementiser '*to*'.

(5:199) a) *Deshalb können Sie Ihre Maus ignorieren, wenn Sie möchten.*

(GTRANS_INSTR_009)¹²²

¹²¹ also known as Pennsylvania Dutch

¹²² The Duden lists a similar usage of '*möchten*' as a main verb: '*Wenn du willst, können wir gleich gehen.*'

b) *So you can park your mouse if you want to [].*'

(EO_INSTR_009)

Some early modals in child language as part of the process of gradual acquisition of the verbal system can be seen as unanalysed lexical entries. Single modals in child language with no subordinate infinitive have been described as a non-adult-like use of modal verbs as main verbs in German, but also in Dutch (e.g. Behrens, 1993: 65; Jonkers and Ruigendijk, 2013: 117). The GECCo corpus does not include child language, apart from some passages in fictional dialogues where modal verbs are used in a way that would be considered incorrect in standard German, for example as linking elements between a subject and an object (5:200).

(5:200) *Darf ich heute früh Puddingsuppe?* (GO_FICTION_010)

Advertisement texts and other persuasive texts are another area where modal verbs are used in a non-standard way as deviations from the norm to attract the attention of the reader or listener (e.g. in tourism texts: *'Wir können auch Berge und Strand!'*¹²³ or in a political text: *'Wir haben gezeigt: Wir können Krise. Wir können Aufschwung. Und wir können Solidarität'*¹²⁴).

¹²³ <http://www.saarland.de/73677.htm> [last checked 08/02/2016]

¹²⁴ https://www.igmetall.de/0165209_600_800_berlinbs_2010_11_c55c89dfd9c6098cc58b79224e8d886243052883.pdf [last checked 08/02/2016]

In general, we assume that English has more and longer verb phrases with more possibilities than German to leave out verb phrase elements. German will probably have fewer verbal ellipses than English that function as cohesive devices. It is possible that a higher number of German verbal ellipses are cases of gapping, a type of ellipsis that usually occurs across two different clauses within the same sentence.

This section described the possibilities for lexical verb ellipsis taking into account the different structures of verb phrases that allow for omissions of the main verb and syntactic restrictions on ellipsis. We assume that verbal ellipsis as a cohesive device most frequently involves the omission of the lexical verb.

5.2.5 Verbal ellipses vs. verbal substitution

Annotators may face potential difficulties when distinguishing verbal substitution from lexical verb ellipsis involving a form of the verb ‘do’ in the remnant due to the close similarity of both structures. Sometimes the distinction between ‘do’ as a main verb, a proform or an auxiliary may not seem straightforward. Halliday and Hasan (1976: 112ff.) distinguish between ‘do’ as lexical verb, general verb, pro-verb and verbal operator. When a form of ‘do’ is used as a verbal substitute it replaces a lexical verb (5:201).

(5:201) *Has the plane landed? Yes, it has [].* (lexical verb ellipsis)

substitution: *Yes, it has done.* / full form: *Yes, it has landed.*

Clear cases of verbal substitution are infinite forms of ‘do’ that replace the lexical verb so that it is not possible to add a lexical verb after ‘do’. ‘Do’ in those cases acts anaphorically as a semantically relatively empty and unspecific main verb to substitute for an antecedent lexical verb (5:202-204).

(5:202) *She looks no different than she used to do.*

(5:203) *He left earlier than he might have done had this injury not occurred.*

(5:204) *I can't hurry more than I am doing.*

Clear cases of verbal ellipsis, on the other hand, involving a form of the verb 'do' (cf. also Halliday and Hasan's discussion of the distinction between 'do' as substitute and verbal operator, 1976: 127) are structures where a lexical verb from the co-text may be added and where 'do' is used as an operator or auxiliary as in (5:205-206) where 'do' does not replace the lexical verb, but is part of the negation or the question.

(5:205) *'I hear things about the ship' – 'The ship keeps changing names. You hear that?'¹²⁵, – 'No, I don't [].'¹²⁶*
(EO_FICTION_003)

(5:206) *Monsieur Cendrier said we should call the police. – And did you []?'¹²⁷*

¹²⁵ Text-type specific, informal auxiliary drop, cf. Chapter 6.5

¹²⁶ This is one of the examples where an object that depends on the verb phrase is omitted in addition to the lexical verb. It falls under verbal ellipsis in my analysis as it can be seen as an omission within the extended verb phrase (cf. pp. 177; 207).

¹²⁷ Example from the book *Malediction* (2012) by Sally Spedding, p. 106, <https://books.google.de/books?id=jo0UYZcp78cC&pg=PA106> [last checked 04/02/2016]

In cases of verbal substitution by a finite form of ‘do’, the lexical verb is replaced by a proform and not omitted, but for some annotators this may seem ambiguous in certain cases (for instance after emphatic ‘do’ [5:207] or in comparative constructions [5:208], where some may see ‘do’ as an elliptic support verb, a proform or a main verb).¹²⁸ They have been marked as ‘problematic’ in the annotation.

(5:207) *This agreement applies in full to the civil aviation industry. The EU itself has explicitly agreed that it does.* (EO_SPEECH_009)

(5:208) *The bit string of addressing information in the RNA gives this system the power of tremendous precision, just as the binary bit strings used by digital computers do.* (EO_POPSCI_004)

It is not the case in our corpus data, but it is theoretically possible that lexical verb ellipsis may have multiple antecedents or a ‘complex antecedent’ (Hobbs and Kehler, 1997: 401), which makes it reasonable to assume that the underlying structures in ellipsis sites do not necessarily have to mirror the structure of antecedent verbs but can also be a verbal substitute or only a specific part that several antecedent phrases have in common:

¹²⁸ The annotation of verbal substitution had been annotated previous to the ellipsis discussion and annotation in the GECCo project, involving semi-automatic procedures.

(5:209) *Mary wants to go to Spain and Fred wants to go to Peru, but because of limited resources, only one of them can [do it / go (there) / go to Spain or go to Peru].*¹²⁹

Historically, ‘do’ has always been a rather general verb that can have a range of meanings as a main verb (e.g. ‘make’, ‘carry out’, ‘act’, ‘handle’, ‘put’, cf. OED entry¹³⁰) or that can be used as a substitute for another verb. As a substitute, similarly as in ‘do so’, it usually, in older stages of the language, had an agentivity restriction with regard to its antecedent, describing the activities of agents and being incompatible with nonstative antecedent verbs, cf. also Houser’s (2010a) remarks on the construction ‘*swa don*’ in which the predecessor of ‘do’ occurring with the deictic-anaphoric manner adverbial ‘*swa*’ was a full verb having the sense ‘to act in such a manner’. The fact that, in Modern English, it is still rarely used with nonstative or eventive antecedents shows that even as a proform or substitute it does have some semantic content. Halliday and Hasan (1976:126) described ‘do’ as a pro-verb in ‘do the same’ and ‘do so’ which means that ‘do’ should be analysed as a generic element – a proform – that can replace a constituent. Nevertheless, the verb ‘do’ on its own does not replace the verb. In (5:210), we cannot use the verb ‘*feeds*’ instead of ‘*does*’. Only ‘so’ or the whole construction ‘do so’ can be seen as replacements of the antecedent verb phrase.

¹²⁹ Webber (1978) described such examples as Inferentially-Determined Antecedents where the antecedent of the ellipsis is derived inferentially by a contextual assumption or presupposition.

¹³⁰ <http://www.oed.com/view/Entry/56228> [last checked 12/02/2016]

(5:210) *Melvyn feeds the cattle, but Louie never does so.*

The syntactic properties of this example are slightly different from ‘*Melvyn feeds the cattle, and so does Louie*’ where the anaphor can be licensed by auxiliary verbs (e.g. ‘*Melvyn has fed the cattle, and so has Louie*’), cf. Hankamer and Sag (1976: 415f.) and Houser (2010b). Miller (1990: 302) saw ‘*do so*’-substitution related to ‘*do it*’ and ‘*do that*’ where ‘*do*’ clearly is an instance of the main verb ‘*do*’ and not of the auxiliary. ‘*Das/dies/es tun*’ is also the German equivalent for ‘*do so*’ in many translated texts, although, as ‘*tun*’ is often avoided in non-translated German texts, it does sometimes sound like a relatively literal translation (e.g. 5:211-212).

(5:211) a) *We are certainly in a position to do so.* (EO_ESSAY_015)

b) *Wir sind sicherlich in einer Position, dies zu tun.*

(GTRANS_ESSAY_015)

(5:212) a) *Unfortunately, the regime shows no inclination to do so thus*

far. (EO_SPEECH_006)

b) *Leider zeigt das Regime bisher keine Neigung, dies zu tun.*

(GTRANS_SPEECH_006)

The status of German ‘*tun*’ / ‘*machen*’ as a semantically bleached pro-

verb and the category of verbal substitution are less clear in German. Only in a few exceptional cases – such as *fronted* infinitives that emphasise verbs at the beginning of a clause – does German use periphrastic constructions with an auxiliary ‘*tun*’ (5:213).

(5:213) *Im Zentrum der Großstadt arbeitet man nur noch, wohnen tut man am Rande der Großstadt* (in analogy with other sentence-initial infinitives followed by finite modal verbs: ‘*Wohnen kann / sollte man am Rande der Großstadt.*’)

Periphrastic ‘*do*’ has also been characterised as a means to avoid difficult verb inflection in German, which has been suggested as an explanation for its frequent occurrence in child language (Erben, 1969: 46, Bonitz, 2012: 17). ‘*Tun*’-periphrasis is a salient feature of German which is frequently encountered in modern German dialects or older stages of German, but which is virtually ungrammatical in standard modern German and ‘*tun*’ in general is often avoided for stylistic reasons (cf. Langer, 2001, on the stigmatization of ‘*tun*’ from Early New High German onwards). Therefore, the use of verbal ellipses in this context is mainly restricted to informal conversations, dialect syntax and a few specific syntactic patterns (Zifonun et al., 1997: 1253; Schwarz, 2004).¹³¹

¹³¹ This may be on the rise in German due to intensive language contact with English and numerous translations of English dialogic texts where these constructions frequently occur (films, series, novels). The lemma ‘*tun*’ is not used very frequently in our written original and translation corpus data. It is more frequent in the spoken data, and most frequently in idiomatic expressions (‘*etwas*’ / ‘*mit etwas zu tun haben*’, ‘*sich*

Nevertheless, it would be relatively unusual to copy an antecedent verb phrase into an ellipsis site at the beginning of a sentence.¹³² Those cases can rather be analysed as examples of topic drop where a pronoun such as ‘*das / dies*’ may be inserted into the ellipsis site. Theoretically, there are several possibilities to complete (5:214-215) by adding syntactic material at the beginning or the end of the German clause and it is not exactly the same grammatical construction as in English (cf. the discussion of examples 5:180-181 above).

(5:214) *Tut das Gerät noch funktionieren? – Nein, [] tut es nicht mehr.*

(While the question in this example is an informal and marginalised use of standard German ‘*Funktioniert das Gerät noch?*’, the answer is grammatically acceptable in standard German)

(5:215) *Geh und sag doch der Dame, dass wir Banditen sind! – [] Tüt ich ja gern, aber das glaubt die mir nie!*¹³³

A contrastive overview on cohesive substitution in English and German

schwer tun’), used as an infinitive, sometimes as 3rd person singular form and rarely in any other form.

¹³² The same actually applies to other cases of German lexical verb ellipsis as in ‘*Ja, [] werde ich gleich [].*’ / ‘*Nein, [] habe ich nicht [].*’ where a main verb or ‘*tun*’ as a general replacement verb would be added at the end of the sentence, but not in the first position.

¹³³ Example from a German translation of a French comic: ‘Lucky Luke 21: Vetterwirtschaft’ by Goscinny / Morris, p. 12, <https://books.google.de/books?id=jjM4CgAAQBAJ&pg=PA12> [last checked 04/02/2016]

is given in Kunz and Steiner (2013) where it is stated that English verbal substitution does not have a parallel structure in German and either corresponds to various types of ellipsis or more specific verbs for lexical cohesion.

5.2.6 The distinction between verbal and clausal ellipses

In both languages, all types of verbal ellipsis can also involve omissions of other syntactically obligatory constituents external to the verb itself which can contribute to the syntactical incompleteness of the remnant structure (Halliday and Hasan, 1976: 197; Schmolz, 2015: 106). The omission of verb phrase elements can, for instance, co-occur with the omission of the subject in operator ellipsis (which is also called ‘ellipsis from the left’ in English) or the object in some cases of lexical verb ellipsis (also called ‘ellipsis from the right’ in English). English verbal ellipses usually have a fixed position in the sentence. For German, we can also generally consider operator ellipsis as ellipsis ‘from the left’ and lexical verb ellipsis as ellipsis ‘from the right’. German has a relatively fixed position of verbs as well; however, in German it is sometimes less clear than in English whether to insert elided material before or after the ellipsis remnant as there can be more patterns of possible underlying structures due to a freer word order.

As explained in Chapter 4.2, the line drawn by Halliday and Hasan between verbal and clausal ellipsis is not very sharp as they stated that it is possible to look at verbal ellipsis from another angle, taking the clause as the point of departure, and to interpret it as a clausal ellipsis. Even if verbal and clausal ellipsis are relatively similar concepts for Halliday and Hasan, it makes sense to keep verbal and clausal ellipsis separate in the annotation scheme according to the criteria described below. An analogous distinction, though based on slightly different criteria, had been drawn

between verbal and clausal substitution in Halliday and Hasan (1976) and in the annotation scheme from previous studies in the context of the GECCo project (Kunz and Steiner, 2013). The corpus annotation will also distinguish between verbal and clausal ellipsis for project-wide terminological consistency.

Nevertheless, due to the conceptual similarity of verbal and clausal ellipses and in anticipation of the frequency results obtained and described in Chapter 9, it should already be pointed out at this stage that we will later subsume these – after a separate annotation – under a more general category for the quantitative analysis. After the separate annotation of verbal and clausal ellipses, we noticed that there are rather low frequencies of both verbal and clausal ellipses that are used as cohesive devices in our data. Although verbal and clausal ellipsis in general are rather broad categories, in many corpus texts a separate analysis of these categories would leave us with an outweighing amount of zero elements in the tables presented in Annex 1. If we want to keep these conceptually similar categories separate in the quantitative analysis, we should obtain a table of the corpus results that would be a sparse matrix.¹³⁴ The number of non-zero elements in the table has been considerably increased when I finally grouped verbal and clausal ellipses under a catch-all heading in the analysis. Nevertheless, much more detailed information on the respective subtype of each example has been collected according to the criteria set

¹³⁴ A sparse matrix is a matrix with a large number of zero elements where the number of zero elements can be much larger than the number of non-zero elements.

out in the annotation scheme so it is possible to analyse these structures and their subtypes separately if one wants to do so.

In SFL terms, a clause unifies different metafunctional strands of meaning as the realization of a message (textual), a move (proposition / proposal; interpersonal) and a figure (experiential), cf. Matthiessen and Teruya (2010: 71). The information unit is a unit that is parallel to the clause and the units that it consists of (ibid.: 27).

Several researchers, e.g. Givón (1979) identified the clause (verb plus any arguments) as the minimal unit of discourse. Haspelmath (2010b: 697) defined the clause with regard to cross-linguistic comparability as ‘an expression that contains one predicate and potentially at least some of its arguments and that can be independently negated’. General remarks on clause structures in English and German can be found in Emonds (2007) and Fischer (2013). Apart from finite clauses, English and German use several types of non-finite clauses such as infinitive clauses, participle clauses or non-finite clausal complements of certain verbs (cf. Huddleston and Pullum, 2002, Chapter 14 and Zifonun et al. 1997: 1375, ff. 2158ff.). We assume that clausal ellipsis typically involves the omission of elements from a finite clause and only in a few cases, for instance in some patterns of sluicing, from a non-finite clause.

Clausal ellipsis, a term popularised by Halliday and Hasan (1976), is the omission of a part of a clause, which can involve one or several constituents. What sets the concept of clausal ellipsis apart from clausal substitution (cf. Chapter 5.3.6) is that in clausal substitution an entire clause is replaced by a substitute, but in clausal ellipsis one or several constituents are omitted from a clause, while at least one constituent is left

as an ellipsis remnant. Clausal ellipsis can occur within an independent clause as in adjacency ellipsis or in a dependent clause, for instance in sluicing.

The subtypes of clausal ellipsis that Halliday and Hasan suggested are total or partial propositional ellipses and total or partial modal ellipses (1976: 196, 335) as Halliday and Hasan analysed clauses in terms of a two-part structure (ibid.: 197) consisting of a ‘modal’ and a ‘propositional’ element. The modal element is understood as a unit of the subject and the finite element of the verbal group and the propositional element as another unit is the remainder of the verbal group and any complements or adjuncts. According to Halliday and Hasan, ‘modal ellipsis is associated with a context where there is no choice of mood in the clause’¹³⁶ and ‘propositional ellipsis is associated with those instances where the mood and the polarity are the principal components of the message’ (ibid.: 198). We do not consider it the most plausible taxonomic classification to describe the different types of omissions from clauses by referring to the complete or underlying structure of clauses by only two elements, one consisting of both the subject and the finite verb and another one containing the rest of the clause. We will therefore not use the terms modal and propositional ellipsis to describe subtypes of clausal ellipsis.

As pointed out in the previous chapter, the line drawn between verbal and clausal ellipsis by Halliday and Hasan is not very sharp when they

¹³⁶ ‘Mood, the choice of declarative, interrogative, imperative and their subcategories, is the realization of speech function, and is expressed by the Modal element.’ (Halliday and Hasan, 1976: 198).

state that it is possible to look at verbal ellipsis ‘from another angle, taking the clause as the point of departure’ (1976: 197). It has already been explained in the previous section that even if verbal and clausal ellipsis are relatively similar concepts for Halliday and Hasan, it makes sense to keep verbal and clausal ellipsis separate in the annotation scheme. Nevertheless, for the reasons set forth above, we will later subsume these – after a separate annotation – under a more general category.

Halliday and Hasan gave a nonexclusive list of examples and subtypes of clausal ellipsis to illustrate the phenomenon instead of providing a precise conceptual definition of clausal ellipsis (e.g. ellipsis in question-answer pairs and other rejoinder sequences). As the classifications used for the annotation should not overlap, and the aim is to place all cases found in the GECCo corpus clearly in only one category, clausal ellipsis is defined here as omissions of constituents that have not yet been covered under nominal and verbal ellipsis.

The most frequent type of non-clausal units and sentence fragments are those that do not particularly refer to a specific textual antecedent but that can be understood using knowledge about the situation, the text type or the world in general. Quite often such fragments cannot be used as a linking element in a text at all and they are not always the result of an omission. Many of these will not fall under clausal ellipsis in the annotation, but under specific fragment categories which will be discussed in Chapter 6.

5.3.2 Adjacency ellipses

Clausal ellipses that are used endophorically as cohesive devices can typically be found as adjacency ellipsis in dialogic interaction in various contexts, e.g. as short answers in question-answer pairs (answers to wh-questions, alternative questions and yes-no questions), elliptical follow-up questions, corrections, confirmations, reduced recapitulatory echo questions, echo exclamations or elliptical echo utterances indicating sarcasm or disbelief (5:216-221, cf. also Quirk et al., 1972: 408ff¹³⁷ and Reich, 2002, 2003, 2007 on ellipses in question-answer sequences).

(5:216) *A: Wouter's farm, along with one other, had been reassigned.*

B: Reassigned? (EO_FICTION_004, reduced recapitulatory echo question, yes-no questions that repeats part of a preceding utterance)

(5:217) *A: You could actually try on some of the shoes and the scarves*

and bits and pieces, which was quite interesting. B: From different historical periods? A: From different eras.

(EO_INTERVIEW_005, follow-up question and a slightly paraphrased echo utterance as confirmation)

¹³⁷ Quirk et al. seem to consider some of these structures such as echo exclamations as a sentence type of their own.

(5:218) *A: I had a brief stay in Georgia. B: Oh, Georgia, ok.*

(EO_INTERVIEW_011, echo utterance, back-channel signal, expressing astonishment and interest)

(5:219) *“Didn’t you ever have fun on vacations?” “Fun on vacations,” she repeated dully.*¹³⁸ (elliptical echo utterance expressing a bored or sarcastic tone of voice)

(5:220) *A: It was a slightly rough area, but it wasn’t quite as bad as Handsworth which is the sort of village, the town in Reigate. B: In Reigate? A: Sorry, in Birmingham.* (EO_INTERVIEW_001, confirmation recapitulatory echo question and elliptical answer as correction)

(5:221) *A: Yes, I’m from the Yorkshire-Lancashire border, a little village called Earby which is on the Pennine Way. And [break]. B: So in the north of England? A: In the north of England, yes.*
(EO_INTERVIEW_006, follow up question and echo utterance as confirmation)

In our corpus data we also find sequence of several rejoinders (5:222):

¹³⁸ Example taken from the book *Revolutionary Road* (1961), Richard Yates <https://books.google.de/books?id=aun3KJlxUHkC&pg=PA256> [last checked 08/02/2016]

(5:222) *A: Also, dort bin ich geboren. B: Dort bist du geboren. A: Aber nicht von schwäbischen Eltern. B: Nicht von schwäbischen Eltern? A: Nicht von schwäbischen Eltern.*

(GO_INTERVIEW_002)

Clausal ellipsis as in adjacency pairs are generally used very similarly in English in German. In our corpus data, we expect to find adjacency ellipsis typically in question-answer pairs and other adjacency pairs in dialogic passages in the registers FICTION and INTERVIEW, representing one-to-one and face-to-face interaction, and in FORUM (polylogues / group communication)¹³⁹, sometimes also in ACADEMIC which predominantly includes lectures as monologic, one-to-many interaction.¹⁴⁰ We expect adjacency ellipsis to be rare in most written texts in our corpus, apart from fictional texts, and only to occur occasionally, e.g. as answers to rhetorical questions without speaker change in registers such as political speeches or tourism leaflets. Most adjacency ellipses are turn-related or back-channel actions reflecting interpersonal aspects.¹⁴¹ It has been suggested that avoiding clausal ellipses can contribute to altering the degree of the speaker's commitment to a proposition as a syntactically

¹³⁹ cf. also Ricento (1987) on ellipses in multi-party conversation

¹⁴⁰ There are cases in GECCo, albeit very few, where rejoinders as part of the conversation cannot be analysed if they have not been transcribed, for example in the register 'ACADEMIC' if it was not recorded when a person from the audience threw in a remark without speaking into the microphone (e.g. *Hat da jemand einen Vorschlag?... Ja, ganz genau.*)

¹⁴¹ A case on which the literature gives no clear answer on how to analyse it syntactically and that may look like an adjacency ellipsis, another type of fragment or a tag question are split questions (cf. Arregi, (2010), e.g. *'What tree did John plant, an oak?'* as a wh-question part and a tag). Nevertheless, such cases do not occur in our corpus at all and they will not be discussed here in more detail.

complete answer may sound more ‘vehement’ (Wilson, 2000: 148). Clausal ellipsis can be used as signal of involvement and familiarity among speakers. It is then considered as a realisation of ‘positive politeness’ (Otsuki, 2009: 32 referring to the term introduced by Brown and Levinson, 1987). On the other hand, brevity can sometimes be an indicative of a lack of commitment to something or even unfriendliness as it can make an utterance sound evasive and dismissive (Nariyama, 2004). Clausal ellipses in answers to questions have also been treated under the term of ‘fragment answers’ in the literature (Merchant et al., 2013). Merchant (2004) proposed an analysis according to which such structures have unpronounced syntactic structure and are derived by fronting a constituent of the answer that provides the new information to a focus position, with the remainder of the sentence being omitted.

There is a considerable variety of structures that can be omitted in adjacency ellipsis: all types of clause structures where only a contrasting or focused element is left in the remnant structure. Remnant structures of adjacency ellipsis can also be diverse. They often consist of a single phrase, but can also have the form of a dependent clause (5:223).

(5:223) *And why am I mentioning this now? [] Because it highlights the different capabilities, also within Europe.*

(ETTRANS_SPEECH_010)

German in general will probably more often have only one constituent left in adjacency pairs where English can have a similarly short structure or a slightly longer one involving lexical verb ellipsis (5:224).

(5:224) *Who was playing the piano? - Peter []. / Peter was [].*

German adjacency ellipsis remnants sometimes carry more morphological information tying the constituents that are left more explicitly to the syntax of the previous sentence. If the antecedent clause contains a prepositional phrase that is repeated in the elliptical structure, the preposition can often be omitted as well in both English and German. Where nouns as remnants are not morphologically marked as the subject or the object by suffixes determiners or embeddedness in clearly parallel structures, there may be a tendency to include more material, such as prepositions, in the elliptical structure, but our corpus is not sufficiently large to test this hypothesis and to specifically compare noun phrases and prepositional phrases. The stranded prepositions in English in the questions maybe another reason why they are not reiterated in the English elliptical answers of examples (5:225-226).

(5:225) a) *What does it look like? – A tape recorder.*

b) *Wie sieht es aus? – Wie ein Tonbandgerät.*¹⁴²

¹⁴² Example taken from the novel *Ishmael: An Adventure of the Mind and Spirit* (1992) Daniel Quinn, p. 49, <https://books.google.de/books?id=83p-OMrNalYC&pg=PA49> and its German translation, p. 36

(5:226) a) *How many subjects are you failing in? – Four.*

b) *In wie vielen Fächern bist du ungenügend? – In vier.*¹⁴³

<https://books.google.de/books?id=EXTMBQAAQBAJ&pg=PT36> [last checked 09/02/2016]

¹⁴³ Mixed case, cf. 8.5. This example has been taken from J. D. Salinger's *The Catcher in the Rye* (1951), p. 10 and its German translation by H. Böll (1962).

5.3.3 Sluicing

Sluicing is a clausal ellipsis phenomenon in which only the wh-complementiser (sometimes also called a bare wh-operator) of a subordinate clause is left as a remnant structure (5:227), cf. Ross (1969). If the wh-word is the subject of the sluiced clause, the verb phrase and its obligatory and optional complements are omitted; otherwise sluicing is an omission of a clausal structure involving the subject as well. A sluiced clause can also have the form of a to-infinitive clause (5:228).

(5:227) *Max went to the store, but Oscar wondered why [].*

(5:228) *I'm tempted to explain, but I don't know how [].*

(EO_FORUM_005)

Some authors, for example, Merchant (2001) offer a broader definition of sluicing than the one that is used in this study. What is sometimes called 'matrix sluicing' includes all wh-question words ('wh-phrases') even if they do not occur in embedded structures as in (5:229). We will treat such cases as elliptical follow-up questions in adjacency pairs and not as typical cases of sluicing.

(5:229) *Someone called. - Who []?*

Sluicing typically occurs within the subordinate clause as the second part of a clause complex (5:227, 5:228), but it can also occur across sentence boundaries (5:229), where its cohesive function is more pronounced as it does not depend on syntactic relations within the sentence.

(5:230) a) *My grandmother was suspicious of us being friends. She never said why.* (EO_FICTION_001)

b) *Meine Großmutter hat es nicht gern gesehen, daß wir Freundinnen waren Sie hat nie gesagt wieso.*

(GTRANS_FICTION_001)

'*Whether*' (as well as '*if*') as a particular example can only appear in embedded questions and not in an independent yes-no questions. Its analysis as a complementiser or conjunction rather than a wh-adverb is not uncontroversial (Trotta, 2000: 34). It cannot appear on its own as a question word or a sluicing remnant (cf. also Cheng, 1997: 33). In large corpora such as COCA we may find rare counterexamples to this rule, for instance when a wh-word is used in a noun-like way (5:231).

(5:231) *All right, for present purposes, all we need to do is figure out who, or more precisely, whether. We already know what, when, and where. The how and why are incidental as far as Mrs.*

Holland is concerned. (COCA, 2005_FIC_Analog¹⁴⁴)

Neither can the German equivalent ‘*ob*’ appear on its own as a question word nor be used as a sluicing remnant,¹⁴⁵ although it can, like wh-question words, introduce certain independent yes-no questions in German (‘deliberative Fragen’ such as ‘*Ob er Anna noch anruft?*’ / ‘*Wen Otto wohl dieses Jahr einlädt?*’). The *Duden Grammar* (2009: 890) describes such verb-final sentences as questions that syntactically look like subordinated clauses, but that are independent clausal units according to their function, punctuation and intonation, cf. also Zifonun et al. (1997: 684ff.). English has an equivalent to such indirect questions only in examples of old-fashioned and formulaic syntax as in the standard opening sentence of the British Prime Minister’s Question time (‘*If he will list his official engagements for today?*’) which reflects British politeness and formality conventions. In general, it can be noted that clauses introduced by a wh-word which can also function as a conjunction cannot be sluiced.

In German, a sluiced wh-phrase bears the same case that it would in a non-elliptical wh-clause (5:232). This is an additional reason for the assumption that the elliptical clause has an underlying complete structure.

¹⁴⁴ <http://corpus.byu.edu/coca/> [last checked 10/02/2016]

¹⁴⁵ Grammatically marginal: ‘*Ich weiß nicht, ob [], aber er könnte Folgendes gemeint haben.*’ (example from: <http://e-flux.com/aup/project/grace-schwindt-interviews-with-germans/> last checked 25/04/2016)

(5:232) *Er will jemandem schmeicheln, aber sie wissen nicht wem [].*

he wants someone_DAT flatter but they know not who_DAT

‘He wants to flatter someone, but they don't know who/?whom/[].’

A special case of sluicing is multiple sluicing where the second remnant wh-phrase is required to be prepositional phrase in English (5:233).

(5:233) *I know that in each instance one of the girls got something for one of the boys. But [which] [for which]? (example from Bolinger, 1978: 109)*

Nishigauchi (1998) and Lasnik (2013) discussed whether this can be analysed as a kind of gapping construction. As multiple sluicing is very rare, like multiple wh-constructions in general, it is non-existent in our corpus data. Sluicing is an anaphoric relation that occurs across clause boundaries. In the analysis of this study, it falls under non-cohesive ellipsis if it occurs within a clause complex as it is restricted by certain syntactic principles and it can be seen primarily as a structural and not a cohesive relation. Sluicing is one of the ellipsis subtypes that have been discussed in great detail in the theoretical literature on ellipsis, but that are relatively rare in our data.

5.3.4 Omissions of predicative expressions

Halliday and Hasan (1976) did not particularly mention the possibility of omitting predicative expressions that serve a cohesive function in a text. Such omissions can be treated as a specific subtype of clausal ellipsis as they are omissions of a constituent that neither fall under nominal nor verbal ellipsis. In our English corpus data, there are anaphoric ellipses of predicate adjectives (5:234), nominals (5:235) and adverbs (5:236). Ellipses of predicative prepositional phrases almost never occur in the data.

(5:234) *She was always talking about who was good-looking and who wasn't [].* (EO_FICTION_001).

(5:235) *Your questions seem to be stemming from assumptions that there is AN answer to the questions. There isn't [].* (EO_FORUM_005)

(5:236) *In practical terms as I say, folk are here because they need to be [].* (EO_INTERVIEW_003)

Although Halliday and Hasan (1976: 203) claimed that it is not possible in cohesive ellipsis to omit single elements from the structure of the clause as they would have to be expressed by a reference item, the omission of predicative expressions seems to be a case of such an omission which

takes the form of the deletion of a single element. In our data, this type of ellipsis typically takes place within clause complexes, but some examples stretch across sentence boundaries. Similar structures in German, on the other hand, usually involve the insertion of a reference item, e.g. a pronoun replacing the predicative expression (5:237b; 5:238b), or they involve a different reduction strategy where more material than the predicative expression is omitted (5:239b).

(5:237) a) *That is not to say that the disputes that do arise are not important. They are [].* (EO_SPEECH_014)

b) *Das heißt nicht, dass die auftretenden Differenzen nicht von Bedeutung sind. Sie sind es.* (GTRANS_SPEECH_014)¹⁴⁶

(5:238) a) *Its institutions are not the natural embodiment of human nature but its aspirations certainly are [].* (EO_SPEECH_012)

b) *Ihre Institutionen sind nicht die natürliche Verkörperung der menschlichen Natur, das Trachten nach Demokratie ist es aber mit Sicherheit.* (GTRANS_SPEECH_012)

¹⁴⁶ What may seem like an ellipsis of a predicative expressions in German in an alternative structure for this example ('[] *Sind sie durchaus*') actually falls under 'topic drop' where a pronoun such as 'das / dies' instead may be inserted into the ellipsis site and the sentence can be syntactically completed with a proform regardless of the context, cf. Chapter 6.5).

(5:239) a) *Additional financial burdens are certain, additional jobs are not [].* (ETRANS_ESSAY_019)

b) *Die zusätzlichen Belastungen sind sicher, die zusätzlichen Jobs nicht.* (GO_ESSAY_019)

Omissions of predicative expressions can sometimes be found in English within a sentence in a comparative construction (5:240a). German mainly uses phrasal comparatives or a pronoun as a reference item which we do not consider to be elliptical (5:240b). For phrasal comparatives, we favour an analysis of a non-clausal structure over an ellipsis analysis in a covert clausal structure (cf. Chapter 5.2.4).

(5:240) a) *You are as much responsible as we are [].*

b) *Ihr seid genauso verantwortlich wie wir.*

In a way, the structure of omissions of predicative expressions resembles that of lexical verb ellipsis after the auxiliary 'be', but they are omissions of a noun phrases, adjectives etc. after 'be' as copula verb. If a predicative expression is elided in addition to an omission of a part of the verb phrase, it will be treated as verbal ellipsis, otherwise as clausal ellipsis.

5.3.5 Mixed cases

If nominal ellipsis co-occurs with verbal or clausal ellipsis (5:241), it has been annotated under a specific category ‘MIXED’. We can either consider it as one specific subtype of ellipsis or two co-occurring ellipses which will make a slight difference for a quantitative analysis.

(5:241) *Four Oysters followed them, and yet another four []*.¹⁴⁷

In co-occurrence of nominal ellipsis with verbal or clausal ellipsis, syntactic material can often be inserted before and after the ellipsis remnant (5:242).

(5:242) *How many slices do you want? – [] Two []*.

(GTRANS_FICTION_008)

Cases where several subtypes of ellipsis co-occur in one remnant structure in this way are sometimes found in question-answer sequences (5:243), but typically in gapping structures in both English and German 0 to keep the second conjunct as short as possible.

(5:243) *Is there quite a high demand then for care nowadays? – [] A*

huge and growing []. (EO_INTERVIEW_003)¹⁴⁸

¹⁴⁷ Halliday & Hasan (1976: 148) mention this as an example of nominal ellipsis although other constituents have been left out as well. It may also be argued in this specific example that it is a case of extraposition of an element from a coordinated structure.

(5:244) a) *Das erste Postulat kennzeichnet das Ziel, das zweite [] den*

Weg, der zu diesem Ziel führt. (GO_ESSAY_001)

b) *Prosperity for all and prosperity by competition are inseparably linked; the first postulate denotes the goal, the second*

[] the road that leads to this goal. (ETRANS_ESSAY_001)

Nominal ellipsis and gapping sometimes co-occur if the first conjunct had a nominal ellipsis of a noun phrase that is contrasted with a noun phrase from the second conjunct (5:245) and (5:55) above.

(5:245) *Wir stehen für die nächsten fünfzig Jahre vor zwei großen*

Herausforderungen. Die eine [] nenne ich die Wiedererfindung

der Demokratie und die andere [] die Neuausrichtung des

technischen Fortschritts. (GO_ESSAY_021)

During the annotation process, mixed cases turned out to be rare so that it has been decided against contrasting them with the other ellipsis subtypes as a separate category.¹⁴⁹

¹⁴⁸ An example from spoken language which is grammatically awkward with an ellipsis remnant consisting of an indefinite article and two adjectives without a noun. Similarly, one usually would not say: *‘There is a huge and growing [].’*

¹⁴⁹ I will consider these rare cases twice in my lists of elliptical structures in the table of Annex 1: a) as nominal ellipsis, b) as verbal/clausal ellipsis. In order to keep the annotation scheme consistent, gapping contexts which frequently contain mixed cases of ellipsis will be put under non-cohesive structures in the category of verbal/clausal ellipsis and cross-clausal nominal ellipsis in mixed cases will be treated as cohesive devices. If mixed cases occur in adjacency pairs, they will fall under cohesive nominal and cohesive verbal/clausal ellipsis.

5.3.6 The difference between clausal substitution and clausal ellipsis

In analogy with the discussion concerning ellipsis and substitution of verb phrase elements (Chapter 5.2.5), the conceptual difference between clausal ellipsis and clausal substitution is established by the presence or absence of a substitute element. A clausal substitute is a place-holding element showing where a clause has been omitted¹⁵⁰ and indicating what its grammatical function would be (Halliday and Matthiessen, 2013: 635). Here again, annotators may face potential difficulties when distinguishing clausal substitution from clausal ellipsis due to its apparent conceptual similarity.

There are only a few specific constructions of clausal substitution in English and a great variety of constructions involving clausal ellipsis. This is due to the fact that clausal substitution is defined as the presupposition of an entire clause (Halliday and Hasan, 1976: 130) while clausal ellipses are not defined as the omission of entire clausal structures, but as omissions of constituents from a clause.

According to Halliday and Hasan (*ibid.*: 130ff.), the word used as a substitute for whole clauses in English is 'so'. A postverbal propositional 'so' can substitute reported clauses, beliefs or assumptions (5:246-247). A preposed 'so' carries stress (5:248-249).

(5:246) *Does that mean he's not going to lose his job? - I believe so.*

¹⁵⁰ or *replaced*, to be more precise

(5:247) *The tuna has been caught without harm to dolphins. It says **so** on the packaging.*

(5:248) *His lawyer had tried, or **so** he said.*

(5:249) *Rage is revivifying, or at least **so** it seems.¹⁵¹*

There is a possibility of overlap where ‘*so*’ is substituting for an attribute or for a fact (5:250).

(5:250) *Is the mango ripe. It seems **so**.* (from Halliday and Hasan, 1976: 134)

Some dictionaries explain this use of ‘*so*’ simply as an adverb with the meaning of ‘MENTIONED EARLIER’. It does not necessarily replace a clause; it can also be used to avoid the repetition of phrases or adjectives that have been mentioned earlier or to state that a situation or a fact mentioned earlier is correct or certainly true.¹⁵² It has a similar function to that of a demonstrative pronoun or other reference items. According to the OED it can assume the function of an object and pass into the sense of

¹⁵¹ Example taken from the book *Group Psychology and Political Theory* (1994) by C. Fred Alford, p.33

<https://books.google.de/books?id=bifM3CY-6fgC&pg=PA33> [last checked 12/02/2016]

¹⁵² <http://dictionary.cambridge.org/dictionary/british/so> [last checked 12/02/2016]

'that'.¹⁵³

The negative form of the clausal substitute according to Halliday and Hasan is 'not' as in (5:251-53).

(5:251) *Descartes walks into a bar. The bartender walks up to him and says, "Would you care for a drink?" Descartes replied, "I think not." and disappears.*¹⁵⁴

(5:252) *Ought we to declare our winnings? It says not.* (example from Halliday and Hasan, 1976: 133)

(5:253) *Has everyone gone home. I hope not.* (ibid)

In my opinion, such cases may also be analysed as ellipsis (e.g. '*I hope [that] not [everyone has gone home]*'). Regardless of the analysis of 'not' as a substitute or an ellipsis remnant, the complete sentence would rather involve changing the sentence structure (e.g. '*I don't hope that everyone has gone home*').

German has no exact equivalent of clausal substitution as a specific grammatical category, as is the case with other substitute elements as proforms in English for nominal or verbal substitution. There are some structures with polarity markers or particles carrying stress that are

¹⁵³ cf. OED entry, sense I 2.a. <http://www.oed.com/view/Entry/183635> [last checked 12/02/2016]

¹⁵⁴ <https://ma.tt/2003/11/descartes-joke/> [last checked 12/02/2016]

grammatically similar to the function of a clausal substitute, but they can also sometimes be analysed as being elliptical (5:254).

(5:254) *Denkst du, es wird regnen?*

- *Ich denke ja [/] / Ich denke schon [?].*¹⁵⁵

- *Ich denke nein [/] / Ich denke nicht [].*

Examples such as (5:251-5:254) are rare in our corpus data, so that, with regard to the data, the question of the difference between clausal substitution and clausal ellipsis may at first sight seem like a theoretical or hypothetical problem. Nevertheless, a structure that does occur several times and brings us to question Halliday and Hasan's definition of clausal substitution is the 'substitution' of conditional clauses (Halliday and Hasan, 1976: 134). Especially in the negative form (5:255), these structures that were claimed to be cases of clausal substitution where 'not / nicht' stands for a clause which is understood from the co-text can equally be interpreted as ellipsis. Additionally, they seem to have become fixed expressions¹⁵⁶ meaning 'otherwise / anderenfalls'.

¹⁵⁵ There are several ways of formulating an answer to the question in (5:254), e.g. with a pronoun reference item: 'Das hoffe ich.' / 'Ich hoffe es.' If 'so' is used, it is fronted, but not stressed and typically co-occurs with a modal particle: 'So denke ich zumindest.'

¹⁵⁶ similarly to certain other structures involving 'if' where the subject and the verb 'be' are omitted: 'if [you are] in doubt', 'if [it is] necessary'

(5:255) a) *That she'd come soon, or **if not** she'd send her husband.*

(EO_FICTION_008)

b) *Daß sie mich bald besuchen würde, und **falls nicht**, würde sie ihren Mann schicken.* (GTRANS_FICTION_008)

We tend to analyse such structures as being anaphoric and elliptical, but due to varying opinions in the literature, they have been marked as problematic in the annotation. Clausal ellipsis in conditional clauses in German sometimes only have the negation marker as remnant left while the remnant structure in English can be longer as in (5:256b) which involves the omission of a predicative expression.

(5:256) a) *Und bezüglich der Off-Shore-Steuer oasen werden wir bald erfahren, ob die geänderte Gesetzgebung im Einklang mit der WTO steht - wenn [] nicht [], werden wir im Frühling oder Sommer ein schwerwiegendes Problem zu bewältigen haben.*

(GTRANS_SPEECH_009)

b) *And on the FSC, we will soon know whether the replacement legislation is WTO compliant -- if it isn't [] we will have a major item to deal with this spring or summer.* (EO_SPEECH_009)

As has been argued above, we will subsume all cases of verbal and clausal ellipsis – after a separate annotation – under a more general category (verbal/clausal ellipses) for the quantitative analysis.

6. Other fragments and reductions

6.1 Overview

Certain types of fragments and non-sentential units do not necessarily involve an omission and should not be confused with ellipsis. Additionally, some reduction strategies and types of incomplete structures are different from ellipses that have the potential to be used as a cohesive device. The ellipsis annotation scheme that has been developed by the author of this thesis within the context of the GECCo research project on German English contrasts in cohesion has a strong focus on elliptical structures that are potentially or actually used as cohesive devices. These categories have been described in great detail and considerable theoretical depth in the previous chapters and have been illustrated with numerous examples.

This chapter describes certain types of fragments and reduction strategies as syntactic phenomena that have sometimes been called ‘ellipsis’ in the literature. I view these structures as being conceptually different from those elliptical structures that have the potential to be used as cohesive devices. In the annotation scheme, they will therefore fall under the umbrella of fragments which includes several subtypes. In addressing fragments, similarly as with respect to ellipses, we will not be concerned with purely semantic implications (‘implikativ’ / ‘einbegreifend’ / ‘mitenthaltend’ / ‘mitmeinend’, cf. Polenz 2008: 25) that

are not reflected in incomplete syntax or non-sentential structures. The results of the annotation of those additional categories can be used, for instance, to make comparisons between incomplete structures ('elliptisch' / 'auslassend' / 'lückenhaft', *ibid.*) and other cases of non-standard syntax or other means of '*Sprachökonomie*' and condensed language ('komprimierend' / 'kompakt' / 'kondensierend' / 'verdichtend', *ibid.*). The overall syntactic or fragmentary nature of texts and registers and the tendency towards syntactic standardisation are other factors that could be assessed. As these fragments are peripheral aspects for the discussion of ellipsis and cohesion and have been annotated mainly for comparative purposes and to clearly distinguish them from cohesive and non-cohesive ellipses, we will describe them in slightly less detail than the other categories from the annotation scheme. They can be summarised in a succinct overview and grouped into certain macro-categories. These macro-categories actually consist of more fine-grained micro-categories, but for our purposes, it will not be necessary to specify the annotation categories in more detail than we have done.

In addition to the ellipsis categories explained in the previous chapters, the following macro-categories of other fragments and reductions have been included in the debate about ellipsis and the annotation process:

- text-type-specific fragments
- sentence splits
- short yes / no replies
- non-clausal units
- other

They may superficially look very similar to the ellipsis categories described in the previous chapters. However, there are various reasons for treating them as specific separate structures. Those who would like to employ a broader definition of ellipsis can tailor the annotation results to their conceptualisation by adding specific categories of fragments or reductions to the analysis of ellipses. Fragments in general have, for instance, been described as a case of clausal ellipsis by Merchant (2004) and Weir (2014) who assume that all types of fragments contain underlying clausal structures. Therefore, it would be possible to subsume selected fragment categories from the annotation under a more general label of clausal ellipsis if one wanted to follow Merchant's and Weir's approach. This would presumably lead to a very high number of clausal ellipses where non-cohesive non-clausal units would outnumber anaphoric ellipses, but as noted above, the focus of this study is on elliptical structures with a potential to be used as cohesive devices.

6.2 Sentence splits

Sentence splits are the result of the isolation or dislocation of sentential elements, e.g. phrases or dependent clauses that have become disconnected from a preceding main clause and punctuated as if they are independent clauses or separate statements (6:1).

(6:1) *You can even stream your songs to the dorm room next door. Or down the hall.* (EO_WEB_005)

(6:2) *Auch die Politik kann der Herausforderung nur gerecht werden, wenn sie Wandel als kontinuierliche Aufgabe begreift. Wenn sie Deutschland für begabte Unternehmer attraktiv macht.*
(GO_ESSAY_002)

They are usually labelled by the broader concept of sentence fragments in prescriptive grammar books, style guides for writers or other pedagogical material and have been described as incorrect usage of punctuation or as grammatical errors in standard written English and German, based on the assumption that such constructions cannot stand alone. Sentence splits are sometimes used intentionally as staccato sentences to capture the immediacy of oral narration or to create unexpected linguistic contrasts in advertisement texts or fictional texts (6:3).

(6:3) *Sie standen sich im Hemd gegenüber. Nachts. Um halb drei. In der Küche.*¹⁵⁷

An irregular punctuation mark can also signal a speaking pause, a pause for effect, emphasis, or reflection, which is a typical device in political speeches and political essays in our data. Probably German – due to its freer word order – makes use of these extrapositions more often than English and places dislocated sentence elements outside the sentence boundary to signal a pause, to emphasise certain constituents or parts of complex sentences or to make a prepared text sound more natural. By dislocation, sentential elements can be shifted to the beginning or the end of a sentence or, as demonstrated by example (6:3), they can become even more isolated from the rest of a statement by adding a punctuation mark that normally separates different sentences. German particularly uses ‘exbraciation’ (Ausklammerung) as a stylistic device as the German verb phrase is ordered according to the sentence brace (‘Satzklammer’) rule. Elements that are usually placed in the inner field (‘Mittelfeld’) can become postponed until the end of the sentence or become more isolated by a punctuation mark. One reason for using an ‘exbraciation’ can be the close contact of a dislocated element to the following sentence, but other reasons have been suggested as well: ‘überfülltes Klammerfeld [...], zu

¹⁵⁷ Example taken from the short story ‘Das Brot’ (1946) by Wolfgang Borchert, in *Das Gesamtwerk von Wolfgang Borchert*
<https://books.google.de/books?id=3HJqAgAAQBAJ&pg=PT272> [last checked 18/02/2016]

schwacher Klammerrand [...], Streben nach sachlicher und emotionaler Hervorhebung' (Sommerfeldt et al., 2001: 249f.). In the course of its development from Old English to Modern English, English has become a 'non-embracing' language with no verb-final appearance and with very few possibilities to place constituents between elements of a complex verb phrase (Stockwell, 1977: 310). Nevertheless, English sometimes uses extraposition, particularly of clauses to move 'heavy' constituents to the end of the sentence due to the 'principle of end weight'.

Sentence splits display similar structures in English and German (6:4). In the first part of this example, the translator changes the perspective by replacing the pronoun and by using a different tense, but the structure of the isolated structure in the second orthographic sentence follows the German original very closely.

(6:4) a) *Sie isolieren Deutschland in Europa. So wie Sie in der Außen- und Sicherheitspolitik unser Land im Bündnis isolieren wollen.*

(GO_SPEECH_001)

b) *She isolated Germany in Europe. Just as you want to isolate our country in foreign and security policy terms in the Alliance.*

(ETRANS_SPEECH_001)

The following sentence split is another example where German, for stylistic purposes, breaks a relatively long sentence into two parts which become two orthographical sentences, although the second is no complete

sentence in the syntactic sense. The English version of this example uses another stylistic device with a slightly different effect and takes up the subject and the finite verb in a parallel structure to obtain two complete and separate sentences (6:5).

(6:5) *a) Wir müssen ehrlich sein, was die Herausforderungen Europas betrifft. Und auf das hören, was uns die Wähler Europas sagen.*

(GTrans_ESSAY_003)

b) We have to be honest about the challenges facing Europe. And we have to listen to what Europe's voters are telling us.

(EO_ESSAY_003)

Where German uses a sentence split in our data, English sometimes puts the same information between dashes as a variation on parenthesis within the sentence (6:6). In this example, it rather disadvantageously separates the head noun from the rest of the noun phrase. The English translator even uses a second parenthesis within this long parenthesis structure.

(6:6) *a) Es gibt unzählige Beispiele, wo Ideen in Deutschland entstanden sind, die Arbeitsplätze aber anderswo. Zum Beispiel die Braunsche Röhre, Konrad Zuses erster Computer oder ganz aktuell die MP3-Technik. (GO_SPEECH_005)*

b) There are countless cases – cathode-ray tubes, for example, or

Konrad Zuse's first computer or – particularly topical – the MP3 technology – of new ideas that have been developed here in Germany but created jobs elsewhere. (ETTRANS_SPEECH_005)

Dislocations of sentential elements ('Herausstellungsstrukturen', for instance 'lockere Appositionen' / 'nachgestellte Erläuterungen', cf. Altmann, 1981) that have been placed at the end *within* an orthographic sentence in order to introduce an example, a clarification or a list after 'namely', 'for example', 'that is', 'und zwar', 'nämlich', 'also', 'etwa', 'z.B.', 'vor allem', 'insbesondere' etc. are not analysed as fragments in the annotation. Often they also are a type of extraposition of 'heavier' constituents ('Ausrahmung / exbraciation', 'Nachtrag' or 'Absonderung').

Sentence splits are primarily a phenomenon of written language in our corpus data where non-standard punctuation can either be conceived as an offense against generally accepted grammar rules or the norms of a text type or as a deliberate stylistic device that reminds the reader of the spoken word. The annotation of spoken registers in our data is more difficult with regard to this category. Different transcribers had to make decisions on where to put sentence boundaries, which may, to a certain degree be seen as an interpretation of the data (Powers, 2005: 45).¹⁵⁸ In our spoken data,

¹⁵⁸ In our case of the GECCo corpus, punctuation has been added in the transcripts of the spoken corpus registers by different transcribers, for instance by the author of this thesis for the German register of spoken academic texts, by students assistants working in the GECCo research project for some other German spoken data, and by different transcribers from various corpus-linguistic projects as some texts and registers have been taken from existing spoken corpora.

the transcription attempts to record as faithfully as possible what was been said and avoids ‘tidying up’ the language, but punctuation usually follows normal orthographic conventions. In corpora with a higher proportion of spoken texts, sentence splits can be found more easily (e.g. [6:7]).

(6:7) *My first job was helping the keeper. With the pheasants.* (example taken from the British National Corpus, HER S_interview_oral_history)

A sentence split in the annotation can reflect an interruption of the speaker by an utterance from a different speaker (6:8-9).

(6:8) *A: So she started selling islands, and she actually became the first real life millionaire as in US dollars B: Wow. A: About 2 years ago.* (EO_INTERVIEW_004)

(6:9) *A: Aber wir sind halt nicht sicher, ob wir das gut finden. B: Aha. A: Diese Kurzpraktika.* (GO_INTERVIEW_012)

Utterances in fictional texts can be interrupted and seem as if they had been split into two parts by the comment of the narrator. The narrator’s comments as well are interrupted by parts from a fictional dialogue in (6:10).

(6:10) *“Vielleicht hast du recht“, räumt er ein, „und diese Weißerde-Männer bleiben immer hier. [...] Jede Geschichte braucht ihre Zeit.“ Und fügt dann etwas hoffnungsvoller hinzu: „Und muss nicht unbedingt traurig ausgehen.“* (GTrans_FICTION_005)¹⁵⁹

If a constituent or a part of a clause at the end of a sentence is separated from the rest of the sentence by a full stop, a dash or a colon, it falls under sentence split in the corpus annotation. Fragments after dashes or colons are often isolated noun phrases or appositions (6:11-12).

(6:11) *He has made sure that the most elementary of life’s risks do not lead to economic disaster for the individual – a fundamental idea that is still taken up today in all countries that are attempting the transition from an agrarian to an industrialized country.* (ETrans_ESSAY_001)

(6:12) *They installed one of the most sophisticated conflict-management systems in the world: collective bargaining.* (ETrans_ESSAY_001)

¹⁵⁹ The English original had a similar sentence split between the first and second comment of the narrator, but not between the utterances of the spoken monologue: *“You may be right,” he concedes. “These Whitemuds may be here for ever. [...] Every story is long.” And adds, hopefully, “It is not necessarily sad.”*

A constituent or a part of a clause may also be split from the rest of the sentence at the beginning of a structure. It is separated from the following sentence to create a pause, usually by inserting a colon or a dash (6:13-14).

(6:13) *Auch hier zeigt sich: Der Osten ist auf dem richtigen Weg.*

(GO_SPEECH_015)

(6:14) *Und: Das schlägt durch beim Export.* (GO_SPEECH_015)

Some cases of this type of fragments may seem to fall on the borderline between a sentence split through the creative use of punctuation marks – a stylistic device to vary sentence rhythm and sentence length – and a cohesive ellipsis. German in particular also uses the possibility to split phrases and to isolate parts of a phrase that are placed at the end of a sentence (6:15) or into a new orthographic sentence (6:16-17) which may seem like an ellipsis ('gespaltene Konjunkte', cf. Höhle, 1983: 3; Lobin, 1993: 69). This usually happens if the second part of the phrase is longer than the first one or if one wants to emphasise the second part.

(6:15) *Seine Tante hat den Hund gefüttert oder den Kater.* (example from Höhle, 1983: 3)

(6:16) *Unterm Strich hat sich diese Langfrist-Strategie für die meisten bezahlt gemacht, zum Beispiel bei der Erschließung neuer Märkte in Asien und in Mittel- und Osteuropa. Oder bei den ersten Versuchen, im Internet Fuß zu fassen. (GO_ESSAY_007)*

(6:17) *Ist das purer Zufall? Oder gar ein linguistischer Scherz? (GO_ESSAY_021)*

In most cases of sentence splits, more (redundant) syntactic material may theoretically be added in a way that takes up the structure of the previous syntactic environment. However, often there is no need to regard these examples as an actual omission. They can more adequately be considered specifications which are added and which may be integrated into the syntax of the previous sentence.

6.3 Answering particles

In the annotation of the GECCo corpus, short *yes / no* replies, their German equivalents, and similar structures where a particle conveys affirmation or negation are treated as a type of fragment. They are polarity or truth markers that confirm or negate the content of a previous polar question (6:18) or statement (6:19) completely or to a certain degree, similar to nonverbal behaviours such as nodding or shaking one's head.

(6:18) *A: Jetzt muss ich ein bisschen weiter ausholen, ist das okay, oder?*

B: Ja, ja. (GO_INTERVIEW_001)

(6:19) *A: You mentioned there about the need to sell your house. B: Yes.*

(EO_INTERVIEW_013)

In Halliday's theoretical framework, answering particles are cohesive ellipses where the non-repetition of the constituents from a question or statement is seen as an ellipsis of the whole clause (6:20).

(6:20) *You mean you were interested in him as a man in private life –*

Yes, yes [I was interested in him as a man in private life].

(example of 'yes/no ellipsis' from Halliday and Matthiessen, 2013: 636)

Halliday and Matthiessen see ‘yes’ and ‘no’ as elliptical ‘adjuncts of polarity’ with ‘so’ and ‘not’ as corresponding clausal substitutes. The suggested derivation of ‘yes’ as a fusion of earlier form of ‘aye’ and ‘so’ by Halliday and Matthiessen (2013: 637) is not the happiest of etymologies and can be rejected as being phonologically inadequate. The word *gea*, *gēse*, *gyse* (with *g* pronounced as *y*) has existed since Old English and language historians do not derive ‘aye’ from this form or see ‘aye’ as a phonetic variant of ‘yea / gēa’ because the vowels do not match and ‘aye’ has no *y*-sound.¹⁶⁰

In the German literature, answering particles have predominantly been analysed as sentence equivalents (‘Satzäquivalente’), sentence substitutes, pro-sentences or sentence words (‘Wörter mit Satzcharakter’).¹⁶¹ They are assumed to have sentential character (‘Satzwertigkeit’), cf. Waltereit, 2006: 57, although they are fragments. While in the opinion of some grammarians their morphological-syntactic classification as ‘particle’ or

¹⁶⁰ Cf. OED <http://www.oed.com/view/Entry/231637/> <http://www.oed.com/view/Entry/14090> and Anatoly Liberman’s column on word origins: <http://blog.oup.com/2014/12/etymology-affirmations-yes-yea-yeah-yep-aye/#sthash.FQjquxVQ.dpuf> [last checked 19/02/2016]

In fact, Halliday & Matthiessen’s suggested derivation of the *s*-element in ‘yes’ to an earlier form of ‘so’ goes back to a hypothesis from the 19th century which was set forth by B. Schmitz in his book on English grammar (*Englische Grammatik, nebst einer literarischen Einleitung in das Studium der englischen Sprache überhaupt*, Berlin 1853, 3rd ed.) as well as in his *Encyclopädie des philologischen Studiums der neueren Sprachen* (Greifswald 1859, 2nd ed.) and which gradually gained some popularity despite a paucity of evidence to support it. This hypothesis was already at odds with the prevailing views of that time and Schmitz was criticised in academic reviews for this and many other of his remarks on historical aspects as being inadequate (cf. reviews in *Archiv für das Studium der neueren Sprachen und Literaturen* Vol. 26 (1859), p. 408, and Vol. 36 (1864), p. 456ff., cf.

<https://books.google.de/books?id=uz1GAAAacAAJ&pg=PA403> & <https://books.google.de/books?id=cT5GAAAacAAJ&pg=PA456> [last checked 20/02/2016]

¹⁶¹ cf. Helbig and Buscha, 1994: 529ff.; Schachter, 1985: 32 and the *Duden Grammar*, 2009: 1052.

‘adverb’ is unclear, as is their connection to ellipsis (cf. for instance Bussmann’s standard reference work on linguistic terminology, 1996: 1062), there is a strong tendency in recent English literature to argue in favour of an analysis as particles being clausal ellipses, rejecting the analysis as clausal substitutes or independent particles not involving ellipsis. This argumentation is based on the assumption that ‘question meanings are functions that, when applied to the meaning of the answer, yield a proposition’ (Krifka, 2001: 288) and that answers to yes/no-questions are derived by ellipsis from full sentential expressions which are identical with the proposition of the question (Kramer and Rawlins, 2009; Holmberg, 2016). It has been claimed that the assumed elliptical structure of answering particles is very similar to that of affirmative replies in languages where polarity questions are answered by echoing the finite verb of the question as a carrier of polarity (bare verb echo-answers, Holmberg, 2016: 68). It is rather difficult to prove with certainty that an entire unpronounced clause has been elided that is essentially identical to the clause in a previous question if no remnant structure of such a clause is left behind apart from a sentence-peripheral particle. If we assume that answering particles such as ‘yes’ involve ellipsis, it would also be possible to complete the structure by a clausal structure that does not have to be identical to syntactic material from a previous question, e.g. by a general statement of agreement or confirmation (*‘Yes, [that’s true / I agree / you are right / I confirm that this is true / the answer to your question is affirmative, it is as you say, the statement made is correct, the request or*

command will be complied with, etc.]'. In English and German, there are several other ways to give affirmative or negative replies using a one-word sentence or a phrase, e.g. *'Indeed.'* (*'In der Tat'*), *'Definitely!'* (*'Auf jeden Fall!'*), *'Exactly.'* / *'Precisely'* (*'Genau'*), *'And how!'* (*'Und wie!'*) that can either be understood as being elliptical clauses or conventionalised exclamative construction or fixed idiomatic expressions.

Sometimes, the exact function of answering particles is not easy to disambiguate as certain polyfunctional particles can be used as degree particles, discourse particles, adverbs or particles of quantity or degree, of affirmation, negation, doubt or indeterminacy or as exclamations that do not refer to a previous utterance (e.g. *'eben'*, *'vielleicht'*, Hartmann, 1979: 131, Waltereit, 2006: 57). German sentence-initial *'ja'*, for example, is potentially ambiguous between a discourse marker signalling hesitation (= *'Well...'*) and an answering particle. Both in English and German, *'yes'* / *'ja'* can be used as encouragement for the speaker to continue speaking, as a particle in response to a question or, in the form of an exclamation, as an indicator to show that you are very pleased about something that has happened or that you are impatient or irritated about something. It can be used to emphasise what you have just said. Additionally, in English *'yes'* can even be used to disagree with something that somebody has said (e.g. *'I've never met her before.'* – *'Yes, you have.'*)¹⁶²

Particles are pragmatic markers and signals of the speaker's intention or of a relationship between two utterances. They typically occur in spoken

¹⁶² http://www.oxfordlearnersdictionaries.com/definition/english/yes_1 [last checked 21/02/2016]

dialogues, but in the annotation scheme, they do not fall under elliptical structures. Answering particles are a particular form of non-sentential fragments which qualify the truth of the previous utterance. Answering particles can be expected to occur frequently in fictional dialogues in our data and in interviews. FORUM is another dialogic register, but as it represents asynchronous, non-face-to-face communication it will not have many answering particles or other backchanneling devices. To study them in more detail and to also cover a wide range of such structures, one would need a corpus with a higher proportion of spontaneous, dialogic spoken language.

6.4 Non-clausal units

Non-clausal units are a relatively broad category in the annotation scheme. They usually consist of keywords and phrases and can function as minor clause types (Huddleston and Pullum, 2002: 944ff.)¹⁶³, fixed expressions or block language (a term introduced by Straumann, 1935: 21). Non-clausal units include exclamations and exclamative phrases (Huddleston and Pullum, 2002: 918ff., Portner and Zanuttini, 2005), usually in the form of single words, e.g. adverbs or interjections (e.g. ‘*Great!*’, ‘*Wow!*’),¹⁶⁴ noun phrases (‘*Charming people!*’, ‘*What a strange thing for him to say!*’), stand-alone dependent clauses (‘*That it should have come to this!*’) or conditional fragments (‘*If it isn’t my old friend!*’). Among non-clausal units are vocatives (e.g. ‘*Sir!*’), fragments and formulae to express congratulations, excuses, thanks, alarms or warnings, stand-alone discourse markers, formulaic non-sentences, fixed expressions, aphoristic sayings and slogans (e.g. ‘*The sooner, the better.*’ ‘*More haste, less speed*’). Some non-clausal units are language-specific expressions, but in many cases English and German have very similar structures, e.g. (6:21).

¹⁶³ Minor clauses have also been called minor sentences (Bloomfield, 1933: 176) or mini-sentences (cf. Peters, 2004: 286).

¹⁶⁴ Interjections include onomatopoeic words, lexical units and fixed interjectional phrases (cf. also Chapter 2.1 in Stange [2016] on the definition of interjections).

- (6:21) a) *These principles, which guide UPS even today, are summarized by Jim's slogan: best service and lowest rates.* (EO_WEB_009)
- b) *Diese Grundsätze, von denen sich UPS noch heute leiten lässt, finden ihren Ausdruck in dem von Jim Casey geprägten Werbespruch: „Bester Service zu niedrigsten Tarifen.“* (GTRANS_WEB_009)

In fictional texts, non-clausal units are typically used as a description of a character or a setting, similar to the stage directions in a drama (e.g. *‘Ein simpler Kurzfilm: ein Schuft, ein Kerl, ein Schuss.’* [GO_FICTION_001] or in examples [6:22-24]). English and German use relatively similar structures in (6:22-23), but English uses even shorter fragments in (6:22) (EO: *‘Van Wartville. 1693.’* vs. GTRANS: *‘Van Wartville 1693’*). English also uses *-ing*-participles in (6:23-6:24) which either have no equivalent in the German structure or correspond to a participle used as an adjective or a finite verb in a subordinate clause in German.

- (6:22) *Van Wartville. 1693. An uprising. A revolt against Stephanus Rombout Van Wart, First Lord of the Manor.* (EO_FICTION_004)
- b) *Van Wartville 1693. Ein Aufstand. Eine Revolte gegen Stephanus Rombout Van Wart, den ersten Lord des Freiguts.* (GTRANS_FICTION_004)

(6:23) a) *Dieser trockene Ton. Die langsam auslaufende Zeit. Eine Höhe, wo die Luft dünn wird. Das Dunkel, wenn Sie so wollen. Oder die Helligkeit, eine gleißende Helligkeit, dabei stehen Sie nur am Rand.* (GO_FICTION_007)

b) *This dry tone. Time going by so slowly. An altitude where the air is thin. Darkness, if you wish. Or bright, a sparkling brightness, and yet you are standing at the edge.* (ETRANS_FICTION_007)

(6:24) a) *A lake, the frigid edge of it, and a canoe walking upside down, six legs between stones and smooth, bent sand. A line of men following.* (EO_FICTION_005)

b) *Ein See. Am kalten Ufer ein Kanu, das sich kieloben auf sechs Beinen bewegt, zwischen Steinen hindurch und über einen glatten, gebogenen Sandstreifen. Dahinter in einer Reihe Männer.* (GTRANS_FICTION_005)

Non-sentential block language can also be used as a technique in advertisement texts to catch the attention of potential clients, especially when visual elements complete the thought. We assume that such fragments may sometimes originally result from sentential sources, but there are numerous arguments in support of a non-sentential analysis of such utterances (cf. Chapter 3.2). It would be possible to subdivide the cases that we see as non-clausal units into numerous more fine-grained

subtypes, but many of these non-clausal structures are typical for specific registers, language varieties and colloquial language. Our corpus does not cover all these types of language usage. If a very fine-grained conceptual characterisation was chosen for the subtypes of non-clausal units, there would be several types of linguistic phenomena for which we would not find enough examples for a reasonable quantitative analysis.

6.5 Text-type-specific fragments

Text-type-specific fragments are non-clausal units and reduction strategies related to certain text types and not yet covered by the categories described above. Such fragments either predominate in texts that belong to a particular type or they are at least among the distinctive characteristics of a specific text type. It would be possible to subsume many of these text-type-specific fragments under the more general heading of non-clausal units based on their structure and nonsentential form, regardless of the text type in which they occur. One could also assume that, in a certain sense, the distribution and frequency of fragments in a text is always text-type-specific by definition and that all texts are characterised by a distinctive profile of fragment structures which help identify the text type. Fragments, and elliptical structures as well, may indeed be one factor that, apart from other text-internal and text-external criteria, lead us to conclude that a text belongs to a certain text type or register.

We decided to include text-type-specific fragments as a separate category in the classification scheme in order to take the non-sentential nature of certain registers of our corpus into account and to build on the terminology and concepts suggested by grammarians such as Klein (1985) who introduced the term ‘Textsortenellipse’ as well as Quirk et al. (1985: 845-848) and Crystal (2003: 216) who identified reduced registers that exist in many, if not all, languages and described certain forms of fragments in relation to specific text types, e.g. block language in personal

letters, cables, diaries, notices and labels, abbreviated structures, nonsentences and minor sentences in instructional writing, broadcast commentaries, weather forecast, advertisements, informal conversation and newspaper headlines.

The text-type specific usage of fragments in our corpus includes some formulaic nonsententials at the beginning or the end of texts from certain registers, e.g. political speech or letters to shareholders which always start with fragments such as *'Sehr geehrte Damen und Herren'*, *'Dear shareholders'*. Fragment structures in the form of captions under figures or tables in popular science texts fall under this category as well as bulleted items, numbered lists and abbreviated sentences in instruction manuals. In our corpus data, only specific registers have headlines and subheadings, therefore such fragments also fall under text-type specific fragments.

Abbreviated structures, also referred to as 'economy grammar' (Halliday 1967: 116) often only consist of lexical keywords and leave out copula verbs and function words such as articles, auxiliaries, pronouns and conjunctions (e.g. *'CEO Resigns, Names Successor'*, *'More cuts to follow — but zero rate unlikely'*, *'Vor 33 Jahren verschwunden: Neue Spur zu vermisstem Kind'*) due to limited space and the need for condensed, compact language. Grammatically acceptable abbreviated language is restricted to very specific syntactic environments. Article drop in particular has often been described as a phenomenon that occurs in relation to specific registers or in headlines. For a detailed discussion of article omissions, cf. de Lange (2008), who points out that missing articles are

also typical in child speech and in agrammatic speech. Additionally, de Lange pointed out that there are certain cross-linguistic differences with regard to article use and article omissions. In her cross-linguistic database analysis and experiments on Dutch, Italian and German, she finds more omissions in the Germanic languages in both adult and child speech, more omissions in sentence-initial position than in sentence-internal position and more omissions in sentences with a non-finite verb than in sentences with a finite verb (ibid.: 66). The ungrammatical use of article drop leads to defective noun phrases. It can occur as a performance error or telegraphic speech in registers that go deliberately or unconsciously against the norms of a text type. Some examples in spoken language may also be explained by phonologic reduction that results in non-rule based, spontaneous omission in spoken language (6:25-6:27).

(6:25) *[] *Student hat den Unfall überlebt. (GO_ACADEMIC_004)*

(6:26) *[] *Logik ist klar? (GO_ACADEMIC_004)*

(6:27) *... *wo wir Werbung gemacht haben, um [] Bus zu benutzen*
(GO_INTERVIEW_005)

In our corpus data, we find grammatically acceptable fragments and abbreviated structures mainly in instructional manuals, tourism leaflets or websites (6:28-6:30).

(6:28) *Vor Feuchtigkeit, Schmutz, starken Temperaturschwankungen und direkter Sonneneinstrahlung schützen. - Nicht in der Nähe von starken elektromagnetischen Feldern benutzen, also fern halten von Funkanlagen oder Mobiltelefonen. (GO_INSTR_012)*

(6:29) *Cardiff The Capital of Wales (EO_TOU_007)*

(6:30) *Reservierungen und Öffnungs- / Vorstellungszeiten unter der Nummer (202) 357-1500 oder über die Website www.DiscoveryTheater.org (GTRANS_WEB_008)*

Pronouns at the beginning of sentences and a few other unstressed words at the left edge of utterances can be dropped in certain registers in English and German, e.g. diaries, text messages, fictional dialogues or informal conversation (6:31).

(6:31) *[] Ist eine ganz gute Möglichkeit, einen Einblick in aktuelle kognitionspsychologische Forschung zu bekommen. (GO_ACADEMIC_004)*

The auxiliaries 'be' and 'have' can occasionally be dropped in informal English, particularly in yes/no questions, but this is considered wrong in Standard English and we do not find it in our corpus data. Labov (1969) explained these structures as null allomorphs of auxiliaries; they resemble

the copula absence in African American English. They would be typical of the register of colloquial English in nonstandard varieties and are used in fictional dialogues, especially with characters from lower classes (6:32-33), but probably more frequently in original English texts than in translated English texts. German has no syntactic equivalent to this structure.

(6:32) *A fellow rider, a dirty girl with blond dreads, turned around and asked her, “[] You going up to 'Pico?’*¹⁶⁵

(6:33) *“[] You sure about this, Rab?... It don't sound right somehow.”*¹⁶⁶

What we occasionally find in English fictional texts in our data are declarative sentence used as a question without a form of ‘do’. This sentence type reminds the reader of child speech where a question is often indicated only by phonological features and rising intonation. A similar way of question formation exists in German, but only in English does it look like a fragment as questions would typically be formed by adding a form of the verb ‘do’. (6:34)

(6:34) *You hear that? ... You understand?* (EO_FICTION_003)

¹⁶⁵ Example taken from the novel *Purity* (2015) by Jonathan Franzen, p.68, <https://books.google.de/books?id=cT94BwAAQBAJ&pg=PA68> [last checked 10/03/2016]

¹⁶⁶ Example taken from the novel *Redeeming Love* by Francine Rivers, p. 34, <https://books.google.de/books?id=CpNBIVDYB7UC&pg=PA3> [last checked 10/03/2016]

If, however, a constituent has been left out at the beginning of a sentence and refers to a specific textual antecedent which is necessary to comprehend the utterance, it is a type of cohesive clausal ellipsis. All other forms of the above described left-edge deletion as well as topic drop, pro-drop, ‘diary drop’ (cf. Haegeman, 1990, Weir, 2012) fall under text-type-specific fragments if they can be understood when viewed in the context of a given text type without the need to look for an antecedent from a previous sentence. Alternatively, it would be possible to analyse such reductions as situational, non-cohesive clausal ellipsis if no textual antecedent is necessary for the interpretation of the sentence.

Abbreviated structures rarely have a cohesive function. Particularly, if they are related to a certain text type or register, they usually do not rely on other textual elements.

6.6 Other

In the design of the annotation scheme I was aiming to develop a descriptive system to make the whole of our data describable. I wanted to avoid any overlaps between categories or vague definitions. Therefore, the annotation scheme has been presented in a very detailed way and many examples were given to illustrate the annotation categories. Many schemes that have been developed for the annotation of corpus data have inevitable ‘leftovers’ that are subsumed under an unspecific ‘miscellaneous’ category. Initially, we did not want to include a ‘ragbag’ or ‘catch-all’ category into which we simply put all structures that we have not explained or that we are not sure about what they actually are. If such a category is needed, this can be seen as an indication of a premature classification scheme. It is already possible to mark cases which may require further discussion as ‘problematic’ (cf. Chapter 5.2). Nevertheless, we finally decided to add the category ‘Other’ with the intention to check the completeness of the rest of the annotation scheme. This category is intended to cover ellipses that do not fit into any of the other categories and cases that are on the boundary of being omission phenomena and clearly different from the ellipsis and fragment types described above. As the annotation scheme is rather detailed and illustrates the ellipses types clearly, we annotated almost nothing under this category.

There are only a few patterns that fall under this category and they are not very frequent in the data. In the following quantitative analysis, they

will not be included in the number of ellipses or fragments. The purpose of annotating them is to distinguish them from the other structures covered in the annotation scheme. As they share at least some similarities with the phenomena we are interested in, their annotation may be useful for further research in the future.

One type of incomplete structures, for instance, which is different from the ellipsis types described above, is aposiopesis, i.e. when a sentence is deliberately broken off. If an aposiopesis is a stand-alone subordinate clause, it is also called an anapodoton (6:35-36). ‘Anapodoton’ is a figure in which a main clause is suggested by the introduction of a subordinate clause – however, that main clause never occurs.

(6:35) *I don't always use incomplete sentences. But when I do...*

(6:36) *Wer andern eine Grube gräbt...*

Aposiopeses only consisting of one or two function words have been called ‘micro-aposiopeses’ by Imo (2011: 280, for example stand-alone ‘*von daher*’ / ‘*deswegen*’ / ‘*insofern*’ / ‘*obwohl*’ with a hanging implication [cf. also Günthner, 1999; König, 2012]). Raymond (2004) observed that stand-alone ‘*so*’ in English is a similar phenomenon and Mulder and Thompson (2008:185) described ‘final but with “hanging implication”’. Such final stand-alone adverbs or conjunctions will not be treated as ellipsis here. They sometimes occur in our spoken data of the GECCo corpus to fulfil

certain discourse functions in contrast to anacoluthon where an utterance is disrupted and are probably on the way to become grammaticalised discourse markers (cf. Traugott, 1995 and Auer and Günthner, 2005 on the role of the development of discourse markers in grammaticalization theories).

An anacoluthon ('Satzbruch') is an abrupt change in the syntax of a sentence (mainly in spoken language), a figure of speech where a sentence is concluded differently than grammatical rules leads one to expect, in other words in which the expected grammatical sequence is absent (cf. Bussmann, 1996: 53). It can result in a structure that may seem like a fragment and it is considered to be the result of self-correction during unplanned speech. Anacoluthic structures can therefore occur in the spoken corpus data, but also in written data that have not been thoroughly edited, e.g. in the register of FORUM. It has been pointed out that the distinction between certain ellipsis types, aposiopesis and anacoluthon is not always clear-cut, particularly in spoken language. Disruptions of grammatical expectations can be used deliberately or unintentionally. Either they reflect stylistic choice or a grammatical fault. To distinguish between them, cognitive aspects such as intention, utterance planning and anticipation of the interlocutors' knowledge should be taken into account, but they are difficult to pin down empirically (Imo, 2011: 1). Cases of aposiopesis, anapodoton and anacoluthon can be annotated as 'other'. In general, there are only very few examples from the corpus that fall under this category.

Clearly ungrammatical forms of incomplete sentences in spoken language could be performance errors in the corpus texts (6:37-38). They can be categorised under ‘other’, but there are only about a handful of examples. Performance errors would be more typical in spontaneous spoken language where people frequently interrupt each other. Unplanned incompleteness in spoken language as ‘repairs’ when the speaker hesitates or starts anew are not annotated.

(6:37) *What else have we used it ___? We have been using it for our department of computing, various projects, one involving looking at building potential golf areas.* (EO_ACADEMIC_004)

(6:38) *A: Wie bist du darauf gekommen? B: Ja, also, ich bin durch einen Vortrag ___, wir waren damals nach Hamburg gezogen.*
(GO_INTERVIEW_001)

In a few cases structures that may seem like fragments are missing word errors in the corpus texts (6:39-42).

(6:39) *In common with many other countries, ___ is facing two crucial challenges.* (ETrans_SPEECH_002)

(6:40) *So nennen die Wiener die quirlige Uferpromenade der Donau am nördlichen Festland bei der ___.* (GO_TOU_019)

(6:41) *Die Gastfamilie wird versuchen, Ihnen ein Höchstmaß an Aufmerksamkeit ___. (GO_WEB_010)*

(6:42) *Benutzen Sie beispielsweise nur abgeschirmte Schnittstellenkabel zum Anschließen von Computern oder ___. (GO_INSTR_011)*

Question tags as a specific case have been annotated under the category ‘other’. Question tags (e.g. ‘*isn’t it?*’ / ‘*haven’t you?*’ etc.) contain an inverted auxiliary or copula verb and a pronoun and occur with a statement. During GECCo project discussions the question has been raised whether they may be seen as cases of verbal or clausal ellipsis. Some authors have described question tags as ellipsis, e.g. Swan (2005: 179) in his book on English usage which is widely used in foreign language teaching. Nevertheless, question tags are different from short questions in adjacency pairs such as: ‘*I’m going out.*’ – ‘*Who with?*’ It would often be redundant, but still possible, to reply ‘*You’re going out with whom?*’. Question tags are polarity tags that are used to ask for confirmation. They always have the same syntactic form and involve obligatory and not optional deletions. The full construction would be ungrammatical; therefore they are different from the ellipsis types described above. Question tags are pragmatic markers, similar to expressions such as ‘*no?*’ ‘*correct?*’ / ‘*right?*’ which were sometimes used in the corpus texts as well. German mainly uses such invariant all-purpose tags e.g. ‘*nicht wahr?*’ / ‘*oder?*’ / ‘*ne?*’ or a discourse particle (6:43) where English would

have used a question tag.

(6:43) a) *Er war doch in Berlin?*

b) *He was in Berlin, wasn't he?*

Rarely occurring joint utterance construction (also called: ‘collaborative turn sequences’ [Lerner, 1987], ‘joint productions’ [Ferrara, 1992], ‘co-constructions’, ‘shared syntax’ [Helasvuo, 2004] or ‘kollaborative Konstruktionen’ [Günthner, 2013, cf. also Goodwin, 1995, on sentence construction within interaction]) can also be put under ‘other’ if a syntactic unit has been completed jointly by two or more participants in interaction (e.g. A: *Heute ist der?* B: *Erste.*) They appear to be anacolutha or fragments when considered in isolation.

These annotation guidelines for ellipses and fragments that have been developed in the context of the GECCo project and that were presented in Chapters 5-6 have also been summarised in a project report (Menzel, 2014b). Additionally, they have been outlined in an overview document that has been made available to students and researchers who requested it.

7. The relationship between ellipsis and other cohesive devices

Using an endophoric ellipsis is never the only possibility to achieve cohesion in a text. We can see what types of cohesive devices can be chosen as alternatives by looking at the ellipses in our English and German corpus data that have not been translated by similar syntactic patterns in the translations. We expect that many elliptical structures can and will be kept in the translations. However, if this is not possible due to grammatical reasons or if the translator prefers a different structure for stylistic reasons, the repetition of lexical material or the use of pro-forms functioning as substitutes will be the most typical ways to avoid grammatically or stylistically marginal elliptical structures. The decision to use an ellipsis, substitution or lexical cohesion to create textual cohesion influences the structure and density of lexical chains, i.e. lexical sequences of semantically related words in a text. Ellipsis and substitution are usually regarded as grammatical cohesive devices whereas the repetition and reiteration of concepts by lexical items establishing a semantic relation between an anaphoric expression and its antecedent are lexical cohesive devices.

Like ellipsis-antecedent-relations, relations between different lexical items or between lexical items and substitutes can contribute to the cohesiveness and textuality of a text – an effect which can be regarded as less pronounced if these relations occur within the same sentence or even within the same clause or phrase (e.g. synonyms coordinated by *'and'*)

within a phrase). Strictly speaking, only those relations that cannot be explained by grammatical and syntactic relations alone create textual links and count as cohesive phenomena.

Let us briefly compare the different functions of ellipsis, lexical cohesion and substitution. All these devices can function as linking elements in texts. An additional function of ellipsis is to highlight certain constituents that remain in the remnant structure and to avoid a repetition that is considered to be unnecessary or redundant in a text. Using ellipses has an influence on the focus and discourse structure of texts. Ellipsis-antecedent relations usually require certain structural parallelisms – even if they stretch beyond sentence boundaries.

Lexical cohesive devices partly serve other purposes than ellipsis. Synonyms, hypernyms and meronyms as subcategories of lexical cohesive devices contribute to the clarification, explanation and reformulation of passages using a wide-ranging vocabulary and avoiding excessive repetition. These means make text-external and textual semantic relations between lexical items explicit. Almost no two synonyms mean exactly the same thing, and therefore synonyms can reflect different aspects of a phenomenon in a text. English is particularly rich in synonyms so that translators translating into that language have a wide range of terms to choose from. Such a range of choice presents its own hazards to translators because, as pointed out above, most synonyms are only approximate equivalents. General nouns (cf. Halliday and Hasan, 1976: 274), a subtype of hypernyms (e.g. *fact*, *idea*, *thing*, *people* ...) are on the borderline

between grammatical and lexical cohesion. Being rather unspecific hypernyms, they can summarize concepts or express non-specificity or vagueness. The reiteration of the same lexical item, which also falls under lexical cohesion, facilitates comprehension, ensures terminological consistency and avoids ambiguity. Additionally, repetition can be used for metrical reasons or rhetorical effects such as emphasis, or the reinforcement of a previous utterance. Partial repetition is repetition with variation and is used for purposes of simplification or expansion (e.g. also used as hypernym / hyponym: *water* – *groundwater*). Partial repetition combines reiteration with new aspects or strengthens contrast between co-hyponyms or antonyms (e.g. '*Fremdsprache*', '*neue Sprache*' / '*Muttersprache*' [7:1]).

(7:1) *Warum sollten schon Kinder ab drei Jahren eine Fremdsprache lernen? Kinder erlernen eine neue Sprache noch intuitiv und unbewusst, genau wie die Muttersprache.*

Substitutes as another device of grammatical cohesion (cf. Halliday and Hasan, 1976: 91ff.) represent a small group of grammatical elements with little semantic content. They avoid repetition, but do not add much new information and are more salient than lexical devices.

This chapter mainly describes the relationship between ellipses, substitution and lexical cohesion, but it does not address the other two types of cohesive devices: conjunction and co-reference. Conjunctive

relations have different functions. Co-reference items can replace entire phrases, but usually do not play a major role for the replacement of *elliptical* phrases. A pronoun such as ‘*it/es*’, for instance, could replace a complete or an elliptical noun phrase entirely, but it would leave little room for expressing semantic contrasts to an antecedent noun phrase.

We could consider demonstrative pronouns as an exception to this principle if they are contrasted with antecedent noun phrases.¹⁶⁷ This happens in original texts, but we did not find examples in our data where a demonstrative pronoun was used as a translation equivalent for an elliptical noun phrase from an original text. It should be mentioned that certain types of modifiers in ellipsis remnants that combine features of adjectives, determiners and pronouns in one word (e.g. quantifiers) behave similarly to pronouns that can replace entire noun phrases and are on the borderline between ellipsis remnants and co-reference items in syntactically complete phrases.

In the following English example (7:2), the noun ‘language’ is followed by a nominal ellipsis after a numeral.¹⁶⁸ It would equally be possible to use a nominal substitute (‘*their first one*’), a repetition as a lexical cohesive device (‘*their first language*’), a synonym or a paraphrase (‘*their native tongue*’) instead of an ellipsis, but we cannot use only a pronoun such as

¹⁶⁷ Chapter 5 explained why I, in contrast to Halliday and Hasan’s analysis, do not regard English demonstratives and their German equivalents as grammatically defective or incomplete noun phrases, but as replacements of nouns even if they can be followed by an additional noun.

¹⁶⁸ The corpus examples in this chapter do not always involve cross-clausal or cross-sentential ellipses that would have a strong cohesive effect, but they illustrate the general alternatives for ellipsis types that can serve as cohesive devices.

‘it’ or ‘this’ (they either cannot really bear a pitch accent or indicate a prosodically or semantically contrastive focus).

(7:2) a) *Our students take on a new language the same way they did their first [].* (EO_WEB_004)

b) *So lernen die Kunden eine neue Sprache genauso natürlich wie ihre Muttersprache.* (GTRANS_WEB_004)

In (7:2b), the German translator could have followed the syntax of the original by translating the elliptical noun phrase ‘*their first []*’ with ‘*ihre erste []*’. In English, we find many of these adjective/numeral + noun combinations such as ‘*first language*’ which are stronger semantic units than their German adjective/numeral + noun equivalents (cf. the discussion in Chapter 5 and example [5:8]). Several English modifier + noun combinations have separate entries in dictionaries (cf. the entry for ‘*first language*’ in the *Longman Online Dictionary*¹⁶⁹) and are neither clearly a phrase nor clearly a compound. This example semantically corresponds to a German compound in the form of a single orthographic word (‘*Muttersprache*’ or ‘*Erstsprache*’), but not really to a phrase that would be a literal translation (‘*erste Sprache*’). The German nominal phrase in the translation can therefore not be split in the same way as the English one leaving a nominal modifier as remnant. Similar examples of ellipses

¹⁶⁹ <http://www.ldoceonline.com/dictionary/first-language> [09/06/2016]

remnants of English adjective/numeral + noun combinations are (7:3) and (7:4).¹⁷⁰

(7:3) a) *The unit is commanded by a second lieutenant or a first []*.

b) *Die Einheit wird von einem Leutnant oder Oberleutnant geführt.*

(7:4) a) *all countries -- developed and developing []*

(EO_ESSAY_006)

b) *alle Länder -- Industrie- und Entwicklungsländer*

(GTRANS_ESSAY_006)

In these examples, ‘*first lieutenant*’ and ‘*developed countries*’ / ‘*developing countries*’ also belong semantically together – like the German compounds ‘*Oberleutnant*’ and ‘*Industrie- und Entwicklungsländer*’. In German, translation equivalents for these English elliptical noun phrases are compounds that are written as single orthographic words. In (7:3b), a simple nouns (‘*Leutnant*’) and a complex noun (‘*Oberleutnant*’) are contrasted. In (7:4b), a suspensive hyphen in the first part marks a partial omission of the first German compound noun which, in contrast to the English structure, is more clearly a word and less phrase-like (cf. also example [5:119] above in Chapter 5 where ‘*von West-*

¹⁷⁰ These examples have no strong cohesive effect due to the close proximity of antecedent and ellipsis remnant, but they illustrate a general difference between English and German elliptical noun phrases.

nach Ostdeutschland' became *'from western to eastern Germany'* in the translation).

Generally speaking, if an ellipsis is not kept in the translation, the most frequent translation strategy that can be expected is the use of a structure that is more explicit than in the original. Lexical means can be used. In English, substitutes such as *'do'*, *'so'* or *'one(s)'* can also or have to be inserted in English if one does not want to use means of lexical cohesion (7:5)

(7:5) a) ... *die Vorstellung, dass es nicht eine globale Gesellschaft gibt, sondern mindestens zwei miteinander konkurrierende []...*

(GO_POPSCI_002)

b) ... *the idea that there is not a single global society but at least two competing ones ...* (ETRANS_POPSCI_002)

In German, there are no exact equivalents to the English substitutes. Semantically rather empty words such as pronouns or general words (*'so'*, *'tun'*, etc.)¹⁷¹ are used in German too and function similarly to substitutes (7:6). They can be grammatically necessary in German if one does not want to use more specific lexical items.

¹⁷¹ The German forms may look similar to the English substitute items, but they partly serve different purposes and cannot be seen as exact equivalents.

(7:6) a) *Nintendo works hard to protect your privacy – and you should [], too.* (EO_WEB_001)

b) *Nintendo unternimmt eine Menge, um deine Privatsphäre zu schützen – und das solltest du auch tun.* (GTRANS_WEB_001)

Sometimes, more implicit and shorter structures can be used than an elliptical one. An elliptical clause can be shortened so that it will finally contain a slightly different elliptical structure (7:7) or it can be replaced with another, shorter syntactic structure that does not involve any ellipsis (7:8).

(7:7) a) *France and Great Britain have strong security policy capabilities of their own. Germany does not [].*

(EO_SPEECH_010)

b) *Frankreich und Großbritannien verfügen über starke eigene sicherheitspolitische Fähigkeiten, Deutschland [] nicht [].*

(GTRANS_SPEECH_010)

(7:8) a) *Brass bands had not been invented during the time of Giraldus Cambrensis. If they had been [], he certainly would have*

commented on them in detail. (EO_TOU_007)

b) *Blechmusik war zur Zeit des Giraldus Cambrensis noch nicht erfunden worden. Sonst hätte er diese höchstwahrscheinlich*

detailliert beschrieben. (GTRANS_TOU_007)

If an elliptical structure is omitted entirely, it can become a type of fragment such as in (7:9) where only the answering particle is kept in the translation. In this case, only the original text contains a cohesive device. The translation is more implicit with fewer textual links.

(7:9) a) *You hear that? – No, I don't [].* (EO_FICTION_003)

b) *Hast du das gehört? – Nein.* (GTRANS_FICTION_003)

In several cases, a different phrasal or clausal structure than an elliptical one can be chosen that avoids the use of an ellipsis without being considerably more or less explicit (7:10). Nevertheless, avoiding an ellipsis by using a different structure that is neither a substitute, nor a means of lexical cohesion, leads to a text with fewer cohesive devices.

(7:10) a) *Diese Flexibilität wirkt sich für den Investor positiv bei den Arbeitskosten aus, die zwischen 15 bis 40 Prozent unterhalb der westdeutschen [] liegen.* (GO_ESSAY_020)

b) *Investors find this flexibility has a positive effect on labour costs, which can be between 15% and 40% lower than in western Germany.* (ETRANS_ESSAY_020)

This chapter has given an overview on ellipsis and some alternative structures. Additional aspects of the relationship between ellipsis, substitution and lexical cohesion have been addressed in Menzel (2016c).

8. Corpus resources

8.1 Previous corpus-based studies on ellipses

The few previous corpus-based studies on ellipses are based on different theoretical approaches and have to date been rather small-scale. A few corpora exist with certain ellipsis annotations. The annotation guidelines for various dependency treebanks, for instance, include some remarks on ellipses, but these have sometimes been criticised as being underspecified.¹⁷² There are some treebank studies on very specific and well-defined ellipsis types (e.g. Harbusch and Kempen, 2007, and Harbusch, 2011, on clausal coordinate ellipsis in German and Dutch). Some empirical studies on ellipses have their main focus on evaluating the performance of automatic ellipsis resolution systems for machine translation purposes or automatic text analysis and text mining tools (e.g. Yamamoto et al., 1997, Yamamoto and Sumita, 1998). Such empirical studies have focused on a limited, yet heterogeneous selection of ‘ellipsis’ phenomena that are usually relatively easy to spot automatically, e.g. phrase-internal omissions (‘Einsparungen’) where a so-called suspensive hyphen (‘Ergänzungsbindestrich’) marks the omission of a part of a

¹⁷² cf. for instance, the annotation manual for the Prague Dependency Treebank: <https://ufal.mff.cuni.cz/pedt2.0/publications/t-man-en.pdf> [last checked 08/06/2016] or the guidelines for Universal Dependencies, a project that is developing cross-linguistic treebank annotation for many languages (<http://universaldependencies.org/docs/u/dep/remnant.html> [last checked 08/06/2016] -- The ‘Uppsala Group on Ellipsis’ comes to the conclusion that there are not satisfactory solutions to all instances of ellipsis and that ellipsis remains underspecified in the Universal Dependencies guidelines (<http://universaldependencies.org/docs/2015-08-23-uppsala/ellipsis.html> [last checked 08/06/2016])

compound word in German as in *‘Sprach- und Literaturwissenschaft’* (Clematide, 2009).¹⁷³ The resolution of coordinate noun phrases where the common element in all but the last or first of a series of compound words or affixed forms is omitted as in the above cited study by Clematide are an ‘ellipsis’ problem that has been addressed by various computer scientists and computational linguists in studies. The resolution of this type of noun phrases with a lack of duplicate words can also pose difficulties for Named Entity Recognition methods. Buyko et al. (2007), Chae et al. (2014) and Jung et al. (2015) tested methods to identify the underlying structures of named entities in complex noun phrases in the biomedical domain that display different types of backward and forward ‘ellipsis, combinations of both types and “nested” entities within complex noun phrases’ (e.g. *‘human alpha- and beta-globin’*, *‘cytokeratins 8 and 18’*, *‘B and T lymphocyte activation and mitogenesis’* or *‘recombinant human nm23-H1,-H2, mouse nm23-M1, and- M2 proteins’*).

‘Ellipsis’ resolution has also been the object of some studies on dialogue systems, for instance in a study by Streit and Krieger (2004) who were interested in developing a multimodal system that can be used as an appointment management system. If a user of the dialog system utters a sentence like *‘Ich möchte einen Termin eintragen’* (*‘I want to enter an appointment’*), the system will present a new appointment entry. If the user adds something after a pause (e.g. *‘mit Schmid’* [*‘with Schmid’*]) as an

¹⁷³ This example is a case of ‘ellipsis’ that can be found easily in corpus data: ‚Durch den Ergänzungsstrich, der als Teil der graphematischen Erscheinung eines elliptischen Wortes aufgefasst wird, lassen sich Rechtsellipsen in einem Korpus recht einfach einkreisen‘ (Clematide, 2009: 40).

elaboration of the preceding utterance in the form of a fragment, the system has to connect the two utterances in a correct way. In this study the term ‘ellipsis’ was mainly used for non-clausal units and sentence splits (cf. Chapter 6.2), similarly to studies by Johnson (1994) and Fernández and Ginzburg (2002) on non-sentential utterances in dialogic corpus data.

Clarke (2012) looked at a variety of functional-structural ellipsis types in an English corpus of about 125,000 words. Many of the ellipses he described are the result of coordination as in (8:1) and involve subject ellipses (‘Subjektlücken’)¹⁷⁴ in coordinated clauses.

(8:1) *He was neat and tidy and [he was] determined to get forward at every chance.* (Clarke, 2012: 168)¹⁷⁵

Such cases of subject ellipsis are not the main focus of my analysis of ellipses contributing to textual cohesion and their frequencies will not be contrasted with cohesive ellipses in this thesis. Non-repetition of subjects and auxiliary verbs can be considered the norm when clauses with the same subject are coordinated and the repetition of the subject and auxiliary in every position where it is possible within a clause complex would be considered highly redundant (cf. Greenbaum and Nelson, 1999: 123). Although they are a frequent linguistic phenomenon, I decided to exclude them from this study.

¹⁷⁴ Höhle (1983)

¹⁷⁵ Additionally, one could see an ellipsis in *He was neat and [he was] tidy...* as it is theoretically possible to repeat the subject pronoun and the auxiliary here as well.

The examples above illustrate some of the diversity of phenomena that has been analysed in empirical studies on ellipses. Only a few corpus-based studies exist that focus on ellipsis types resembling the categories used in my study. Such studies tend to be restricted to English data and to very specific ellipsis subtypes. They usually cover either certain nominal or verbal ellipses. Günther (2013), for instance, conducted a detailed study on nominal ellipses in English on the basis of examples from the British National Corpus. We will describe her query methods in more detail in Chapter 8.4 as we tried out similar queries in the GECCo corpus to evaluate their applicability for our purposes.

Bos and Spenader (2011) semi-automatically annotated certain patterns of lexical verb ellipsis in the *Wall Street Journal* corpus. Shahabi and Baptista (2012) conducted a corpus-based translation study on English-Persian verb phrase ellipsis. As had already been demonstrated in empirical studies by Hardt (1997) and Nielsen (2003, 2004, 2005), the automatic identification of lexical verb ellipses is difficult and requires detailed manual corrections, which Bos and Spenader's study confirmed.

Miller (2014) worked on an investigation of the use of pseudogapping in the Corpus of Contemporary American English (COCA). He used automatic queries to identify potential pseudogapping contexts. For practical reasons, he had to limit his study to only those pseudogapping cases that have full noun phrases or object personal pronoun as remnants

after an auxiliary or after a form of 'do'.¹⁷⁶ Even these restricted searches led to a lot of 'noise', especially in the case of 'do', so that the relevant examples had to be selected manually in an additional step.¹⁷⁷ Miller finally found only about 1,400 occurrences of pseudogapping in this corpus of 450 million words.¹⁷⁸ As it is often the case with ellipsis contexts, the automatic search for such a very specific construction like pseudogapping characterised by a distinct syntactic pattern still leads to considerable noise so that the actual examples – which are not very numerous – have to be selected manually in a tedious and time-consuming process. Our GECCo corpus is much smaller than the COCA corpus and therefore not really large enough to focus only on one specific subtype of ellipsis in a quantitative analysis.

Nevertheless, the results from previous corpus-based studies highlight the value of corpus data as the patterns and distribution of elliptical structures found in the data often deviate from assumptions and standard examples from the theoretical linguistics literature. Halliday and Hasan, for instance, made several statements about the frequency of certain types of ellipses in English on the basis of personal linguistic judgment. Various quantitative studies based on corpus data seem to contradict certain assumptions by Halliday and Hasan. Moreover, the statements from the theoretical literature remain rather vague on what it actually means if an

¹⁷⁶ e.g. *'It doesn't bother me,' I said. 'Well, it does me,' he growled. / We'll let you know if it deals with the heat and humidity as well as it did the frigid slop.* (examples from Miller, 2014)

¹⁷⁷ The pronoun 'you' had to be excluded entirely as it does not have a distinct object form and consequently led to considerable noise from Subject-Auxiliary Inversion.

¹⁷⁸ Miller used the 1990-2012 version; the current version of COCA is larger as it includes data from 1990-2015, <http://corpus.byu.edu/coca/> [last checked 08/02/2016]

ellipsis type is described as being ‘frequent’ in a language. It has been suggested, for instance, that the omission of the subject and the operator in question-answer sequences where the reply looks like a non-finite verbal group (as in ‘*What should she have done? – Told the police*’) is supposed to be ‘very frequent’ according to Halliday and Hasan (ibid.: 191). We do not expect such structures to occur many times across our corpus registers. It has also been claimed that lexical verb ellipses are found more often between sentences than in coordinated structures (ibid.: 175) and that elliptical verbal groups in which the operator occurs alone are ‘extremely frequent’ (ibid.: 127). However, in a corpus study on lexical verb ellipsis in English, Hardt and Rambow (2001) showed that relations between verbs and textual antecedent only occur occasionally.¹⁷⁹ The mean word distance between an ellipsis and its antecedent in their study was between 6 and 7 and the mean sentential distance was between 0.1 and 0.2.¹⁸⁰ The study showed that this type of ellipsis typically occurs in close proximity to its antecedent and is only used in some cases to establish cohesive relations across sentence boundaries.¹⁸¹

The corpus study by Hardt and Rambow leads us to assume that ellipsis as a cohesive device in general is used occasionally in texts. Ellipses

¹⁷⁹ Hardt and Rambow were interested in variables that correlate with verb phrase ellipsis in order to build algorithms for the generation of ellipses in texts. The distance between the antecedent and the ellipsis site measured in number of sentences and words were found to correlate significantly with the presence of lexical verb ellipsis while aspects such as the length of the antecedent verb phrase had no significant influence.

¹⁸⁰ A value of 0 means that the ellipsis and antecedent verb phrases are in the same sentence.

¹⁸¹ The study by Hardt and Rambow confirmed that distance is not the only factor that affects the decision on using an ellipsis. Syntactic parallelism was confirmed as another factor that contributes to the probability of an ellipsis occurring, for instance, if the adjuncts of the antecedent and candidate verb phrase are the same, lexical verb phrase ellipsis is more likely to happen.

represent grammatically acceptable deviations from standard syntax, but they are mainly used as exceptions in opposition to general norms. Many ellipses will occur at a short distance from their antecedents. Therefore we expect a relatively high proportion of non-cohesive cases among all types of ellipses. This can perhaps be explained by Walker's (1996) argument in favour of the principle of limited attention capacity in processing discourse phenomena. Moreover, recognising omissions as cohesive devices and relating them to textual antecedents may require even more attention than relating overt lexical material to antecedents.

In this section, we provided an overview of previous corpus-based studies on ellipses in order to determine whether certain hypotheses or methods from such studies can also be applied to the annotation or interpretation of our data.

8.2 GECCo – details on corpus design and compilation

The GECCo project and the corpus that has been compiled for the analysis of cohesive devices in English and German have already been mentioned at several points in this thesis. The purpose of this chapter is to describe the corpus and its particular characteristics in more detail. We will address the tools and methods related to the compilation, annotation and retrieval processes. To the best of my knowledge, there are no English or German corpora available so far representing a wide registerial range where cohesive ellipses have been annotated. The GECCo corpus was chosen for this study due to the fact that it was designed for the analysis of all types of cohesive devices. It is probably the largest and most representative bilingual corpus of English and German texts providing comparable data of both written and spoken registers for these languages.

The bilingual, multilevel-annotated GECCo corpus which has been developed as a resource for a contrastive investigation of cohesion consists of ca. 1.69 m. tokens (we chose to measure tokens rather than words as the definition of words is not straightforward). It includes texts and transcriptions from a broad range of written and spoken registers and text types. The written part of the corpus is a comparable and, at the same time, a bi-directional parallel corpus. It contains comparable texts of English and German as well as their sentence-aligned German and English translations. It is comprised of fictional texts (FICTION), political essays (ESSAY), instruction manuals (INSTR), popular science texts (POPSCI), letters to

shareholders (SHARE), prepared speeches (SPEECH), tourism leaflets (TOU) and corporate websites (WEB). The written part of GECCo is a modified version of the former CroCo corpus, which was developed in the framework of the CroCo project on linguistic properties of translations. It has been described in project deliverables¹⁸² and in Hansen-Schirra et al. (2012). GECCo's written part was designed to have an overall size of 1 million words so that each sub-corpus contains approximately 250,000 words, and each register contains roughly 31,250 words. It was compiled in line with Biber's ideas of corpus representativeness (Biber, 1990; 1993) which are widely reported and generally accepted as standards.¹⁸³

The spoken component is a comparable corpus. In an earlier stage of the project, it consisted of academic lectures (ACADEMIC) and interviews (INTERVIEW). The current corpus version includes more spoken registers than before. The recently compiled registers are talk shows (TALKSHOW), internet forums (FORUM), medical consultation (MEDCONSULT) and sermons (SERMON). They have been described in more detail in Kunz et al. (submitted). File headers contain metadata, for instance, data on the speakers, authors and translators and information from a register analysis that has been carried out for each text. Each register contains at least 10 texts and ca. 35,000-40,000 tokens¹⁸⁴ per language and production mode.

The labels ACADEMIC, FORUM, ESSAY, etc. in the metadata of our

¹⁸² http://fr46.uni-saarland.de/croco/deliverable_en.html [last checked 05/05/2016]

¹⁸³ Biber claimed that well-balanced small corpora are capable of covering all linguistic features of a given register. He recommended a minimum number of 10 texts per register with a length of at least 1,000 or 2,000 words.

¹⁸⁴ including punctuation marks

subcorpora are not theoretically well-motivated ‘registers’ in the strict sense of the word,¹⁸⁵ but they are rather socio-cultural intuitive categories. The same applies to our labels ‘spoken’ and ‘written’. There is no binary boundary between spoken and written language. Fictional texts (FICTION), for instance, include both narrative and dialogic passages. Academic lectures (ACADEMIC) are monologic, planned texts where some passages are read out or recited. Most registers in GECCo represent specific public communication contexts and are different from daily conversations in many respects. Our data in the register INTERVIEW come closest to everyday language. All interview texts contain one-to-one, face-to-face conversations about who the interviewees are, what they do for a living and where they live. The corpus texts in the registers of INSTR, POPSCI, SHARE, TOU and WEB were written by experts in a certain field and address heterogeneous groups of non-specialists, but they often assume a certain level of basic knowledge and interest on the topic on the part of the reader. The texts in ACADEMIC, ESSAY and SPEECH also present information in a way that is accessible to a large number of people. These registers contain many elements from the common language and popularise specialised knowledge to a certain degree. The political speeches (SPEECH) in our sample consist of rather formal, carefully prepared text where politicians address both experts and the public so that this register has many of the characteristics of written language *and*

¹⁸⁵ Lee (2001) clarifies and teases apart the terms genre, register, text type, domain, sublanguage, and style in detail. Determining linguistically-based register boundaries will be one of the results of our project.

particular stylistic and syntactic choices used for certain effects in spoken language. The data in FICTION and FORUM represent more heterogeneous registers. In both, a variety of topics is covered depending on the specific text and the texts contain narrative and descriptive passages as well as expressive dialogues. In the translations of the corpus texts, we can sometimes observe a slight shift on the written-spoken continuum. Political speeches, for example, that were officially translated serve journalistic documentation purposes and are less intended to be read out. As a general proposition, it can be stated that the corpus includes some registers that are intermediary registers between written and spoken language.

GECCo is a general corpus which includes a wide variety of texts of high quality with very detailed annotations. The corpus has various annotation levels, for example, information on words, lemmas, parts of speech and the form and function of syntactic constituents. The annotated corpus is available in XML format and can be queried with CQP (Evert, 2005). Additionally, there is an online CQPweb version (<https://fedora.clarin-d.uni-saarland.de/cqpweb/>) as a CLARIN-D web service that is freely available for querying the corpus for research purposes. Via the online interface, the annotated corpus can be queried and the results can be filtered, visualised and exported. Further details on the corpus architecture and annotation levels can be found in Lapshinova-Koltunski et al. (2012), Kunz and Steiner (2013) and Menzel and Lapshinova-Koltunski (2014).

All types of cohesive devices identified by Halliday and Hasan (1976) have been or are being annotated semi-automatically or manually in the corpus. Endophoric ellipses as a type of cohesive devices have received less attention in the first project phase of GECCo than textual realisations of other subtypes of cohesion which turned out to be relatively easy to detect automatically and to disambiguate – such as the phenomenon of substitution for which only a small number of specific function words had to be queried or co-reference and conjunction which have already been conceptually well developed in previous monolingual studies on cohesion. For the present study, all original texts and their sentence-aligned translations from the English and German written corpus registers have been annotated for ellipses. From the spoken component, ACADEMIC and INTERVIEW were annotated for this analysis as these two spoken registers were included in the released version of GECCO at the time of the annotation (cf. Lapshinova-Koltunski, et al. 2012).¹⁸⁶ The spoken part of the corpus has recently been extended to include more registers. New spoken registers were integrated into the corpus after the ellipsis annotation process and are not yet annotated for ellipsis in this analysis. However, in order to have more spoken registers annotated than only two, one of these new registers, namely computer-mediated polylogues from internet forums (FORUM), has been added to the annotated dataset to analyse ellipses. Being a subtype of computer-mediated communication, this register represents a particular communication situation different from

¹⁸⁶ Corpus versions that have been used for this study: GECCO2013, GECCO-SPOKEN2014

the other corpus registers. Although it often reads as if it is spoken, it can also be considered as an intermediary register between written and spoken language. One reason why other new spoken registers, apart from FORUM, have not been added to the ellipsis annotation at this stage is that they seem to be a slightly noisier dataset and potentially less comparable due to reasons that need not be explained here.

Tables 2 and 3 provide an overview of the size and composition of the corpus data that I have used for the analysis.

language	mode	Texts	tokens
EO	spoken	32	121,795
EO	written	110	286,331
ETRANS		121	322,223
GO	spoken	34	125,537
GO	written	121	288,490
GTRANS		110	284,561
total			1,428,937

Table 2: Size of GECCo subcorpora (based on tokenisation in corpus version GECCo2013)

language	mode	register	texts	tokens
EO	written	ESSAY	29	34,998
		FICTION	10	36,996
		INSTR	10	36,167
		POPSCI	11	35,148
		SHARE	13	35,824
		SPEECH	14	35,062
		TOU	11	35,907
	WEB	12	36,119	
	spoken	ACADEMIC	10	40,559
		FORUM	10	43,338
INTERVIEW		12	37,898	
ETRANS		ESSAY	23	42,036
		FICTION	10	40,037
		INSTR	14	39,663
		POPSCI	10	37,878
		SHARE	11	39,511
		SPEECH	18	39,766
		TOU	22	43,675
		WEB	13	39,657
GO	written	ESSAY	23	35,668
		FICTION	10	36,778
		INSTR	14	36,880
		POPSCI	10	36,177
		SHARE	11	35,235
		SPEECH	18	35,399
		TOU	22	36,574
	WEB	13	35,779	
	spoken	ACADEMIC	10	43,703
		FORUM	10	41,636
INTERVIEW		14	40,198	
GTRANS		ESSAY	29	35,345
		FICTION	10	37,652
		INSTR	10	35,820
		POPSCI	11	33,603
		SHARE	13	36,375
		SPEECH	14	35,909
		TOU	11	34,152
	WEB	12	35,705	

Table 3: Size of GECCo registers (based on tokenisation in corpus version GECCo2013)

Figure 7 and Figure 8 visualise the size of the subcorpora and the registers in GECCo.¹⁸⁷

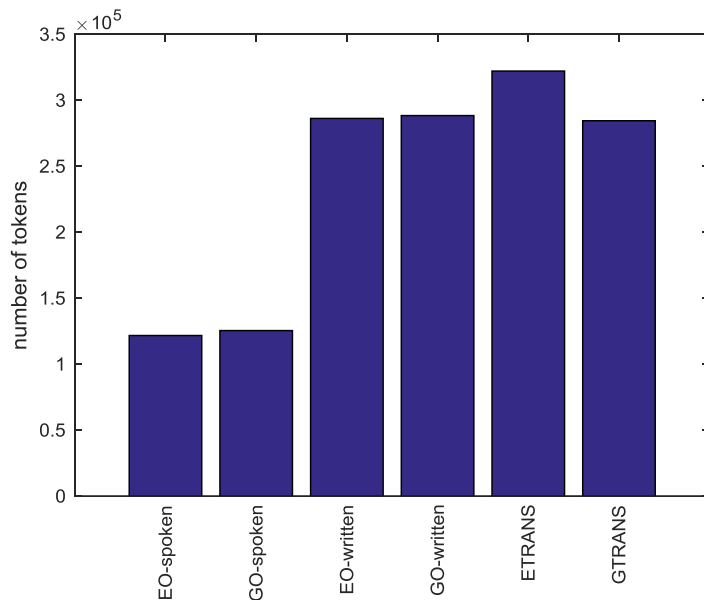


Figure 7: Size of GECCo subcorpora

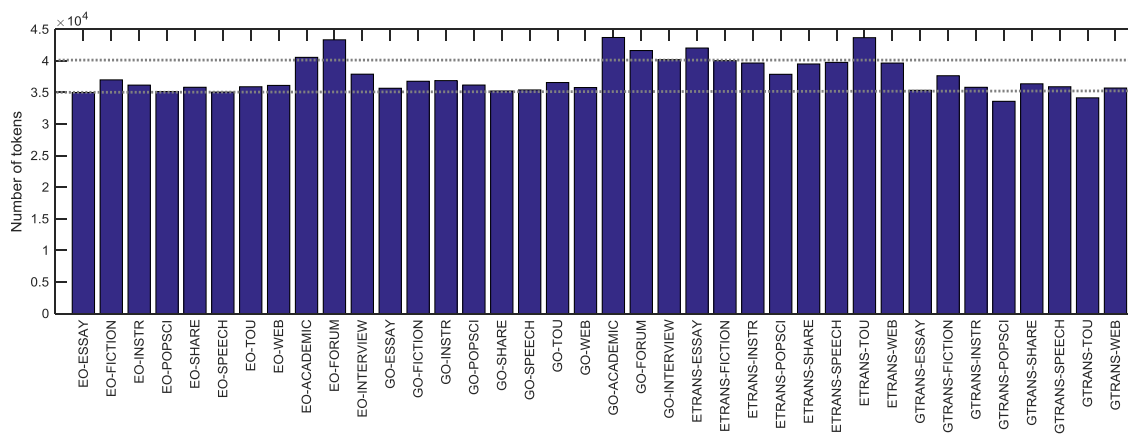


Figure 8: Size of GECCo registers

¹⁸⁷ The plots in this section were produced with MATLAB R2015a, Mathworks (more details on the use of MATLAB in Chapter 9)

These tables and figures on the size of the corpus and its subcorpora and registers show that all registers have roughly the same size in tokens.¹⁸⁸ The spoken registers have a slightly higher number of tokens than the written registers. The English and German spoken subcorpora have a comparable size and include fewer registers than the written subcorpora. Among the written corpus sections, the registers in the English translations subcorpus are longer than the other written subcorpora. These differences with regard to the size of our registers and subcorpora will make it necessary to use normalized frequencies for many calculations in the analysis.

As a bilingual corpus, the GECCo corpus is not as large as some other synchronic, monolingual corpora such as the British National Corpus¹⁸⁹, the DWDS corpora or the IDS corpora such as COSMAS and DEREKO¹⁹⁰, but it can be considered to some degree representative for English and German and for original texts versus translated texts. No individual texts or particular text types in GECCo are overrepresented in a way that would lead to skewed results. The corpus mainly represents standard written and spoken English and German as used in Germany, the United Kingdom and the United States.¹⁹¹

¹⁸⁸ Initially in the process of the corpus compilation, the number of words and not tokens was the criterion for the appropriate text length of the samples. The aim was that each register in the CroCo corpus, GECCo's predecessor, should contain roughly 31,250 words. I think that a cross-linguistic definition of 'word' is problematic for typologically different languages. The number of tokens is a better indication of the size of the texts.

¹⁸⁹ <http://www.natcorp.ox.ac.uk/> [03/06/2016]

¹⁹⁰ <http://www.dwds.de/ressourcen/korpora/> & <http://www1.ids-mannheim.de/direktion/kl/projekte.html?L=0> [03/06/2016]

¹⁹¹ The spoken texts, particularly the German interviews, are characterised by some lexico-grammatical features of local varieties.

Figure 9 visualises the individual text lengths in tokens in all English and German original texts, their subcorpora of written and spoken texts GECCo and in the subcorpora with the English and German translated texts.

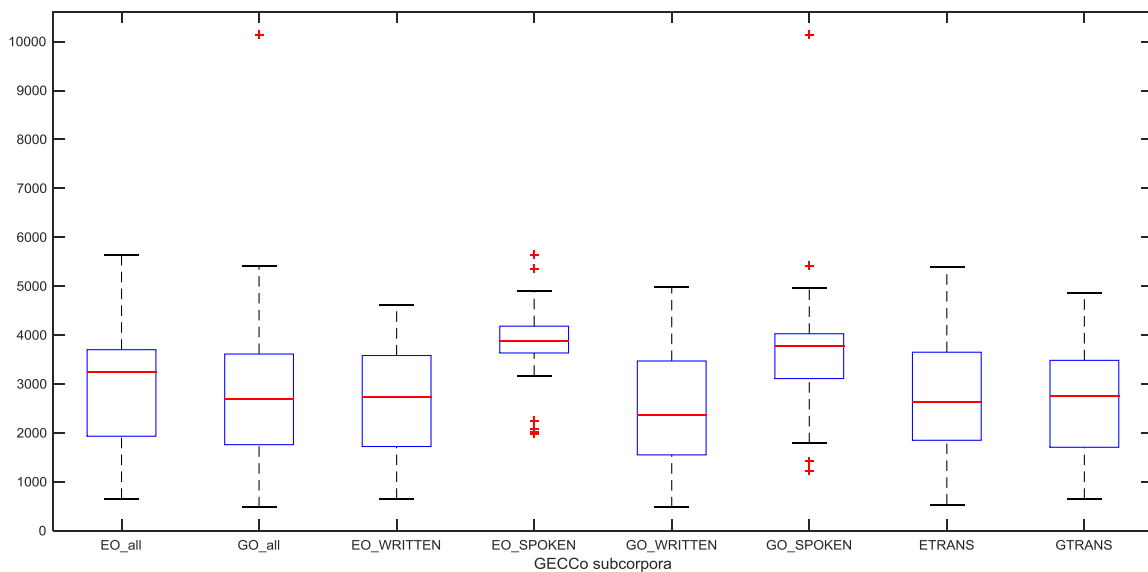


Figure 9: Number of tokens per text in GECCo subcorpora

In Figure 9, we have used a boxplot as visualisation technique. On each box, the central mark is the median, i.e. the point on the measurement scale below which 50% of the texts are located. The edges of the box are the 25th and 75th percentiles. The whiskers extend to the most extreme data points that are not considered as outliers. Outliers are plotted individually using the '+' symbol. MATLAB draws points as outliers on the basis of the interquartile range.¹⁹²

Figure 9 demonstrates that our English and German texts are similar

¹⁹² <http://de.mathworks.com/help/stats/boxplot.html> [last checked 12/05/2016]

with regard to their respective length. The median is only slightly higher for the English original texts. Although some texts are a bit shorter and some are longer, most texts contain between 2,000 and 4,000 tokens. Usually the texts contain at least 1,000 tokens, which we consider a minimum text length that corpus texts should generally have so that ellipses are likely to occur at least once.¹⁹³

At this point, we would like to address the potential limitations of the corpus. We acknowledge that it is possible to see certain limitations in the sampling design that are due to practical and resource constraints. Admittedly, one could say that it is questionable whether a register comparison based on 10 texts per register with about 30,000 words in total is based on sufficient data at all. A variety of statistical studies recommend taking at least 30 samples to quantify mean values. Many statistical tests – such as the Chi-Square Test – need at least 20 or even 30 data points to produce reasonably valid and reliable results (cf. for example McPherson, 1990: 262).¹⁹⁴

¹⁹³ All subcorpora also include a few texts that contain between 4,000 and 6,000 tokens. The spoken texts in the data are usually longer than the written ones. GO_SPOKEN has a very long text with about 10,000 tokens as an outlier. It is not possible to exclude the outliers and unusually long or short texts in our analysis as the corpus is considered to be balanced in its current composition. In general, it would also have been better if texts shorter than 1,000 tokens were avoided entirely in a corpus that is used to investigate ellipses, but at least only very few written texts in the corpus are shorter than 1,000 tokens. We should keep in mind that ellipses will probably occur at least once in most spoken corpus texts, but not necessarily in the written texts, simply because the median length of the spoken texts is higher than that in the written texts.

¹⁹⁴ We sometimes hear the slightly misleading rule of thumb that you need thirty samples for statistical significance. It can be questioned whether 30 is the ‘magic number’ for sample sizes. One argument for the number 30 for a statistical analysis is that the t-distribution becomes a close fit for the normal distribution when the number of samples reaches 30. In fact, the more samples we use for the estimations of population parameters such as the mean or the standard deviation the closer they will be to their true population values. Therefore, we should have a reasonably large sample size (some say at least 30 as a rule of thumb, others would claim that there should be at least 10 data points).

Additionally, sampling has to be entirely random in order to apply statistical tests. The corpus texts represent fairly random text samples, but as the written corpus part is a parallel corpus, only those types of registers and texts were added to the corpus for which professional translations were available. Therefore, the corpus texts usually contain some content and topics that must have been relevant to an international audience or readership at the time. Standard written corpora in general do not contain non-published texts such as private letters, handwritings or texts that are not conveniently accessible. Moreover, random sampling requires a determinable population, but this is difficult with languages in general and with texts that are associated with certain registers. The sampling method that was chosen at the beginning of the corpus compilation was quota sampling, i.e. a selection of texts according to fixed quotas, e.g. one text per author.¹⁹⁵ Quota samples do not strictly represent probability samples that allow for valid generalisations to a larger population. Nevertheless, the sampling methods used for the GECCo corpus approximated random selection and eliminated as many sources of bias as possible.

Another point that needed to be considered in sampling is that heterogeneous registers (e.g. fictional texts or websites) may need different sample techniques than more homogenous and more specific registers (e.g. letters to shareholder). Usually, heterogeneous populations should first be subdivided into more homogenous parts or strata (e.g. FICTION can be divided into mystery novels, thrillers, romances, science-fiction novels,

¹⁹⁵ cf. Deliverable No. 1 'Corpus Design' of CroCo, the predecessor project: http://fr46.uni-saarland.de/croco/corpus_design.pdf, p.5 [last checked 05/05/2016]

etc.) and then each stratum can be sampled independently to obtain a mini-reproduction of the population. In corpus registers with a sample size of 10 texts, it is not reasonable to expect that every part of the population is represented in the sample.¹⁹⁶ The GECCo corpus takes account of the characteristics of the individual corpus registers. The texts in more heterogeneous registers cover a diversity of topics in a way that makes them as comparable as possible across languages.

Text availability is one of the well-known constraints on corpus-building in general. Many corpora are convenience samples (i.e., you use what you can get). The concept of corpus representativeness has been discussed extensively in the literature and it is nearly impossible for any corpus to contain a sufficient number of text categories and sufficiently long texts within each category to fully represent any given language. In the GECCo project, we work with a rather small data set of bilingual data and we are aware of the fact that corpus size has an influence on the level of statistical significance as the probability of rejecting a null hypothesis increases with the sample size. We expect to obtain meaningful results for specific research questions but it will be easier to make statistically valid claims about larger subcorpora of data, e.g. to compare the whole English dataset to the German data, or all originals to the translated texts. Due to the fact that there is no clear-cut boundary between spoken and written language in our corpus and that the size of the spoken data is smaller, this

¹⁹⁶ Moreover, words such as homogenous and heterogeneous are ambiguous as long as the observation scale and the set of constitutive elements taken into consideration are not defined with precision (cf. Gy, 1992: 49).

study will mainly focus on general differences between English and German and between its individual registers – and less on differences between written and spoken language. Neither does this study intend to focus on a detailed comparison *within* the different registers of spoken language, as in the spoken part of our corpus, spontaneous or conversational everyday speech, multi-party conversations and private conversations are slightly underrepresented due to data scarcity and the difficulty to integrate a large amount of such data into a corpus. Another aspect to bear in mind with regard to the written corpus data in GECCo is that it only contains texts that were available in printed or online form at the time of corpus compilation. Surely the corpus should not be seen as a repository of perfect texts.¹⁹⁷ Although there is some variation in the editorial quality of the corpus texts, most of them seem to have been editorially polished for publication.

We intend to obtain certain relevant and meaningful results for register variation and for the differences between written and spoken mode as well, but on the basis of fewer data points these in particular should be viewed with less statistical stringency. Our spoken registers have certain features of prepared or written language and within the written registers there are,

¹⁹⁷ During the annotation process for ellipsis, a few specific corpus texts turned out to involve several flaws that we should keep in mind in a corpus analysis. One translated text in ETRANS (ETTRANS_WEB_013) seems to be of poor quality and it includes several grammatical and orthographic mistakes. It was probably not translated by a English native speaker or by a professional translator. Our GO_FICTION texts and their English translations involve another difficulty for the analysis of discourse structures. In several cases they consist of extracts from different chapters of a novel, probably depending on the availability of free sample translations for these passages at the time of the corpus compilation (GO_ETTRANS_FICTION_002, GO_ETTRANS_FICTION_004, GO_ETTRANS_FICTION_006). In SHARE, sometimes several letters to the shareholders of a company were put in one file although they date from different, consecutive years and now represent one corpus text.

for instance, dialogic passages of fictional characters' spoken language. Cross-registerial and register-internal variation can be evaluated on the basis of the corpus data as well, but we should keep in mind that the individual texts have a reasonable – but not excessive – length and the corpus registers consist of 10 texts or, otherwise, of a higher number of shorter texts. The above mentioned corpus limitations are due to practical reasons and it is easy to find serious limitations in any other general corpus. Let us take the large and highly respected British National Corpus as an example. If we look at the BNC metadata, several questions come to mind, for example why all texts representing the register of SERMON are sermons that were preached by the same person – Reverend Albert Gunter – representing one person's style and the religious practices of one particular Evangelistic fellowship in Essex.¹⁹⁸

The finer details of the conceptual framework for this thesis and my hypotheses on ellipses were formed *following* data collection, which may not seem to be an ideal choice at first sight, but the corpus data were collected in several stages and with several goals in mind where the investigation of ellipses as textual links was only one aspect. Nevertheless, GECCo was designed as a general corpus intended to be used for the analysis of all types of cohesive devices regardless of whether they are rare or frequent. To my knowledge, our corpus is the only existing resource that allows for an investigation of different cohesive phenomena cross-linguistically and across different registers at the same time.

¹⁹⁸ cf. <http://bnc.phon.ox.ac.uk/transcripts-html/> [last checked 08/03/2016]

To sum up this section, it can be said that the GECCo corpus is a useful resource for a wide variety of research purposes and as a reference for translators, language teachers or linguists. The GECCO corpus can be used to generate and test various linguistic hypotheses and to observe linguistic patterns and discourse phenomena such as ellipsis-antecedent relations across languages, registers, modes (written vs. spoken) and production types (original texts vs. translated texts) if we keep the above caveats in mind. For the purposes of this study, a substantial part of the GECCO corpus – nearly 1.5 mio tokens – was manually annotated for different subtypes of ellipses and fragments. We will explain the annotation, retrieval and extraction process in the next section.

8.3 Annotation, retrieval and extraction of ellipses

The annotation scheme for the annotation of ellipses in GECCo is based on the conceptual clarifications presented above and some considerations regarding the annotation process were already touched upon in the previous chapters. This section specifically details the annotation, retrieval and extraction processes of ellipses. The annotation of ellipses has been performed manually with the open source annotation tool MMAX2 (Müller and Strube, 2006). More precisely, the ellipsis remnants and not the omitted elements or the ellipsis sites themselves have been annotated.¹⁹⁹ All annotations were produced by the author of this dissertation according to the GECCo project's agreed annotation guidelines that were developed in the context of the dissertation.

In the context of textual cohesion, we have defined ellipsis as a phenomenon where the remnant of a syntactic omission is left grammatically incomplete and the lexico-grammatical content of the ellipsis site can be recovered from its textual antecedent.²⁰⁰ The following categories are relevant for the annotation of the corpus with regard to (potentially) cohesive ellipses (cf. Figure 10):

¹⁹⁹ A summary of the annotation of ellipses in the GECCo corpus with MMAX2 can also be found in a project document (Menzel, 2014b).

²⁰⁰ or, in a few cases, from a postcedent

- nominal ellipsis
- verbal ellipsis
- clausal ellipsis
- mixed cases (co-occurrence of nominal + verbal or clausal ellipsis)

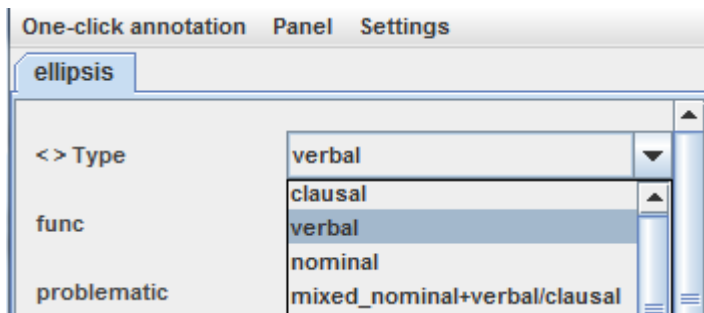


Figure 10: Annotation categories of potentially cohesive ellipsis in MMAX2

Within these categories, there are distinctions between non-cohesive, cohesive and clause-internal structures in the annotation scheme (Figure 11).

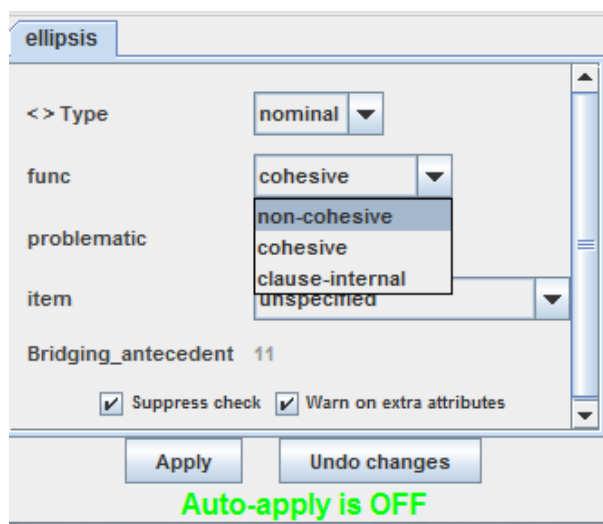


Figure 11: Annotation of function of ellipsis in MMAX2

Nominal, verbal and clausal ellipses are cohesive if they stretch across clauses or sentences and refer to a textual antecedent to establish a textual link. In the annotation, it has to be determined which cases of nominal, verbal and clausal ellipses establish such textual links and are used as text-forming cohesive devices. Ellipses within these categories that formally meet the criteria of incomplete nominal or verbal groups or clauses but, for instance, refer exophorically to the extralinguistic situational context or are rather independent of the context (if they have a standard interpretation as a type of lexicalised ellipsis) fall under non-cohesive ellipses. A specific and relatively frequent subtype of non-cohesive ellipsis are locally bound omissions of nouns referring to antecedents within the same clause or even the same phrase that are mainly the result of intra-clausal grammatical rules. They have been annotated as clause-internal nominal ellipses. Endophoric verbal ellipses also fall under non-cohesive ellipsis here if they are merely the result of coordination or subordination and cannot refer back to textual passages longer than the coordinated or subordinated structure.

The MMAX2 annotation scheme is capable of expressing relations between textual items. Therefore the antecedents of endophoric ellipses have been annotated in MMAX2 as well and a pointer relation links a remnant to its antecedent (Figure 12).

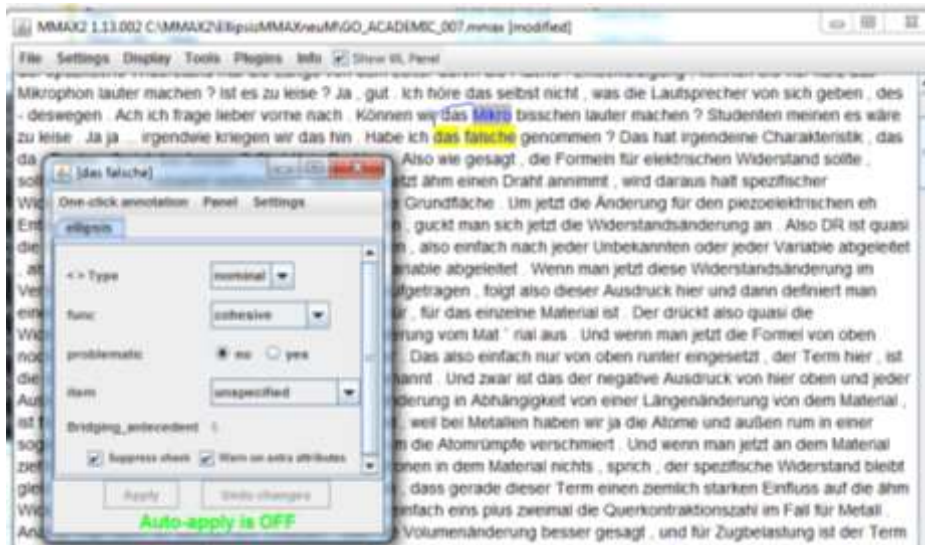


Figure 12: Link between elliptical phrase and antecedent in MMAX2

The antecedent is marked under the level ‘item’ as ‘antecedent’. It can be specified in the annotation whether the antecedent is in the same clause as the ellipsis site or whether there are several possible antecedents as in very few cases the exact textual antecedent cannot be determined due to ambiguity. For ellipses, the level ‘item’ is left as ‘unspecified’ (Figure 13).

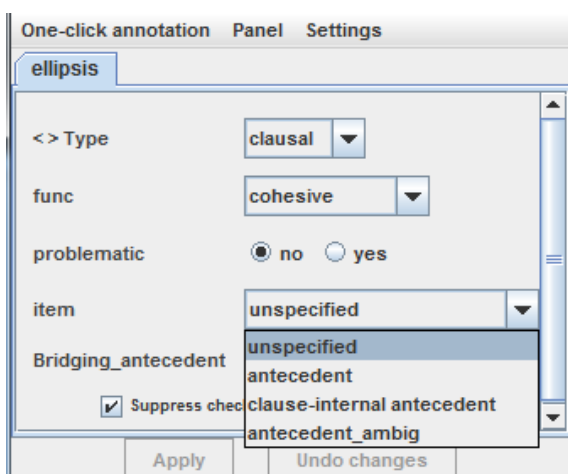


Figure 13: Annotation of elements as ellipses or antecedents

The ellipsis remnant is connected to the nearest antecedent if the structure has been reiterated throughout a text.²⁰¹ If there are chains of ellipses where several ellipses refer to one common antecedent, these ellipses are all linked with the same antecedents for practical reasons. An ellipsis referring to an antecedent within the same clause is marked as clause-internal and its immediate antecedent is marked as a clause-internal antecedent. It is sometimes possible to claim that the ellipsis refers equally to a cross-clausal antecedent and creates a textual link to a context larger than the immediate clause. Figure 14 shows an example of a context displayed in the annotation tool where a nominal ellipsis is anaphorically related to a noun occurring both in the same clause and in one of the previous sentences. Additionally, this example involves a local ambiguity. The context has to be examined in detail as several nouns could be potential antecedents for the ellipsis site, for example the word ‘landmark’ in the previous sentence. After closer reading of the passage, we can infer that ‘cathedral’ has to be the correct antecedent. In the annotation, the ellipsis remnant is linked with the clause-internal antecedent as this noun is more important for the resolution and interpretation of a clause-internal ellipsis than any additional occurrence of that noun in a different sentence.

²⁰¹ Typically, this is an anaphoric relation. Our annotation does not distinguish explicitly between anaphoric and cataphoric relations as cataphoric ellipses are extremely rare in the corpus data or they would rather fall under the category of Right Node Raising. If the element the ellipsis refers to occurs only after the ellipsis site in the text this element will also be marked as an ‘antecedent’ although it is actually a ‘postcedent’.

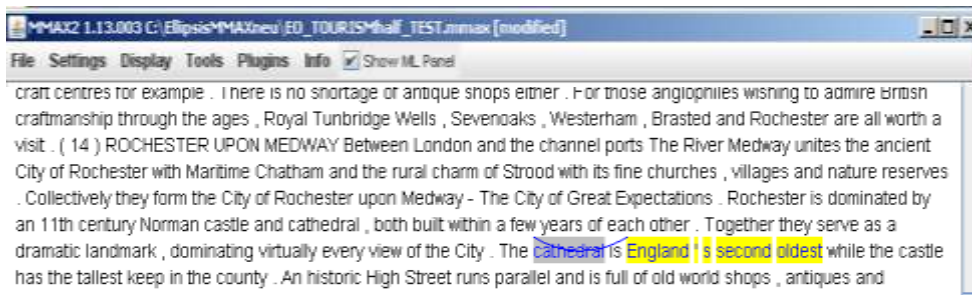


Figure 14: Nominal ellipsis with anaphoric reference to antecedent occurring both in the same clause and in one of the previous sentences

Problematic cases that initially required further discussion, for example, cases that, in the opinion of some, can be considered to be on the borderline to other cohesive devices or structures if there are different opinions in the literature or cases where discussions among project team members over the course of the GECCo project may have revealed slightly divergent opinions on certain patterns with regard to the question of whether they should be analysed as ellipses can be marked as ‘problematic’ (problematic: ‘yes’ / ‘no’, cf. Figure 13).

Furthermore, as has been mentioned before, the cases identified as ellipses have to be distinguished from other types of autonomous non-clausal units, other omission phenomena, text-type-specific block language, abbreviated sentences and non-sentences, sentence splits, short replies where a particle conveys affirmation or negation that may superficially look similar to the ellipsis categories mentioned above but actually need different analyses. Therefore, the following categories as subtypes of other fragments are also annotated for comparative purposes and to distinguish them from ellipses.

- sentence splits
- answering particles
- non-clausal units
- text-type specific fragments
- other

Figure 15 is an extract with examples of a non-clausal unit in a fictional text that uses many verbless clauses. The analysis of this particular example is slightly problematic. Some would consider the exclamation ‘*Schluss!*’ (= ‘*Enough!*’) as a (non-anaphoric) reduced form of a longer structure with various possibilities to add material in order to obtain a complete sentence, e.g. ‘*Jetzt ist Schluss.*’ (= ‘*Enough has been done/said*’). In this context, it rather functions as a quasi-interjection (cf. OED entry on ‘*enough*’).²⁰²

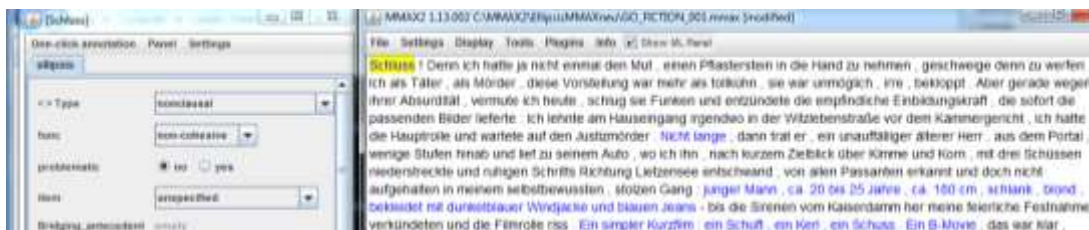


Figure 15: MMAX2 annotation of an example of a non-clausal unit in a fictional text in GECCo

²⁰² <http://www.oed.com/view/Entry/62546> [last checked 22/02/2016] also followed by ‘*of...*’ / (in German: ‘*mit...*’) in interjectional phrases

Due to practical constraints, it was not possible to have more than one annotator for elliptical structures for each text document, as the GECCo project so far had a strong focus on the annotation of other cohesive devices – reference, conjunction, substitution and lexical cohesion where either semi-automatic procedures with manual post-correction by human annotators or more labour-intensive manual annotation was used. In order to evaluate the annotation scheme and to detect and resolve any inconsistencies, the scheme was tested with about twenty students in a seminar on textual cohesion and it achieved high inter-annotator agreement among them for the types of ellipses as defined in the annotation scheme with texts from the GECCo corpus and various other German and English texts.²⁰³ Other students have used the annotation scheme for their term papers. Some of them annotated cohesive ellipsis in a selection of texts from the GECCo corpus without having access to the existing ellipsis annotations, and they by and large produced the same ellipsis annotations as those that are encoded in the corpus. Students who did not use the ellipsis annotation guidelines developed in the framework of this study and who tried to base their annotation of cohesive ellipses on Halliday and Hasan's description and existing German literature on ellipses and textual cohesion always found it extremely difficult to decide

²⁰³ Inter-annotator agreement is often calculated using the kappa statistic (Müller and Strube, 2006). This also takes into account that a certain degree of agreement between the annotations can theoretically be ascribed to chance. If the annotation yields a kappa value below a critical threshold, this would indicate that the phenomenon to be annotated is inherently ambiguous or vague. The original version of kappa is the unweighted kappa coefficient introduced by J. Cohen in 1960. This coefficient can be calculated, for instance, with MATLAB (<http://www.mathworks.com/matlabcentral/fileexchange/15365-cohen-s-kappa> [last checked 26/04/2016]). MMAX2 also contains a special tool for the quantitative comparison of annotations.

on which structures to annotate as cohesive ellipses. In general, the annotators that were given the annotation scheme developed for this study found it considerably less difficult to agree on which items belong to the categories as the guidelines are explicit on the different types of elliptical phenomena and their subtleties and they minimise grey areas. The classifications used for the annotation do not overlap or use gradual categories. The results of the students' annotations confirmed that the ellipsis annotations in the corpus are reproducible. As an additional consistency and quality control check, all annotations were double checked by the author of this dissertation six months after the first annotation was conducted.

After the annotation process, all instances of ellipsis were extracted from the corpus and listed in the order in which they occurred in the corpus. It is a large document; due to space constraints it has not been included as an appendix. The annotated structures can be queried with CQP in the corpus versions which have been made available for online search after free-of-charge registration.²⁰⁴

Table 4 contains examples of queries used for data extraction. For instance, query 1 is used to extract information on non-cohesive nominals

²⁰⁴ In a few cases, if an annotated ellipsis was at the same time part of another annotated fragment structure such as a headline or a split sentence, it was found that the annotated ellipsis was not shown in the CQP query results. MMAX2 allows the annotator to mark elements with different labels in the same file, but apparently, some information from those cases is not displayed in CQP queries after the integration of the information into the corpus architecture. This problem with double-annotated structures only occurs in a few cases, but additionally to the CQP queries in the whole corpus, each MMAX file with the annotations for the individual texts was double checked with MMAX queries to confirm the exact number of ellipses and other fragments per text.

that occur in a particular corpus text.

	Query	Explanation
1	<code><ellipsis>[_ellipsis_type="nominal"&_ellipsis_func!="cohesive"&_text_id="GO_ESSAY_001"&_ellipsis_item="unspecified"]+</ellipsis></code>	- nominal ellipses that are not cohesive and that occur in the text 'GO_ESSAY_001' - the query excludes antecedents of nominal ellipses
2	<code><ellipsis>[_ellipsis_type="verbal"&_ellipsis_func="cohesive"&_ellipsis_item="unspecified"]+</ellipsis></code>	- cohesive verbal ellipsis - no antecedents
3	<code><ellipsis>[_ellipsis_type="clausal"&_ellipsis_func="cohesive"&_ellipsis_item="antecedent"&_text_register="WEB SHARE"]+</ellipsis></code>	- antecedents of cohesive clausal ellipsis in the registers WEB and SHARE

Table 4: Query examples used to extract the categories under analysis

The extracted numeric results are saved in tables for statistical validation.

8.4 Manual vs. automatic annotation procedures for cohesive ellipses

In a first pilot study (Menzel, 2013, 2014a), it was tested for one register in GECCo whether automatic or manual annotation would be a more efficient and reliable method for the annotation of ellipses and other fragments and reductions in the GECCo corpus. Additionally, I tested several query methods in the entire corpus that have been suggested in previous corpus-based studies on ellipses to evaluate their applicability for our purposes. I examined whether potentially cohesive ellipses can be identified with the CQP query language by querying specific part-of-speech patterns and / or certain words that can ‘trigger’ ellipses and whether it is possible to reduce the time needed for ellipsis annotation by combining automatic and manual annotation procedures.

Substitution, co-reference and conjunction had been annotated semi-automatically in the first project phase of GECCo by combining CQP queries and manual analysis. With these cohesive devices, certain specific words and a limited number of structures that can be used in their function as cohesive devices had been queried and pre-annotated automatically and then corrected manually. Human annotators for instance had to check whether these elements were indeed used as cohesive ties or only as structural elements between or within phrases. Some of these semi-automatically annotated devices and their subtypes occur in all or most corpus texts and even if not all instances of a phenomenon such as

conjunctive relations are found by corpus queries, this still gives us a large subset of relevant cases that are sufficient to support or contradict certain hypotheses. These previous studies on substitution, co-reference and conjunction have shown that pre-annotation or pre-tagging of potential candidates for a linguistic phenomenon, in which an existing system automatically annotates texts before it is given to a human annotator who only corrects wrongly labelled elements, simplifies the annotation process in studies on highly frequent elements that are relatively easy to disambiguate. Furthermore, semi-automatic annotation proves useful if no unlabelled data have to be annotated in addition to the pre-tagged items and if the human annotators do not have to check large passages of texts that have not been pre-tagged. Pre-tagging and manual disambiguation has, for instance, worked very well in a study by Skeppstedt (2013) on named entities in English clinical texts, but it would not work in our case where we are interested in a rather rare type of textual relations that involve a great variety of remnant structures in different text types and languages.

If cohesive ellipsis was a highly frequent phenomenon or if our corpus was extremely large, we could decide to restrict the analysis to only those cases that are easy to query as in the study by Günther (2013) mentioned in Chapter 7.1. After having tested several types of corpus queries, we came to the conclusion that only a few subtypes of elliptical structures can be queried with reasonable accuracy with the CQP language in the GECCo corpus, e.g. nominal ellipsis after superlative adjectives, lexical verb

ellipses after modal verbs in sentence-final position or clausal ellipsis in question-answer pairs. If we chose to query only those cases of ellipsis that are easy to query, we would only detect a handful of cases in a corpus of the size of GECCo and miss numerous other omissions of nouns, verbs and parts of clauses in different syntactic environments.

We tried out similar queries in the GECCo corpus to the ones described in Günther's study (2013) on nominal ellipses as we initially assumed that nominal ellipses in general are a type of ellipsis that can be identified most easily with automatic methods by querying noun phrases without head nouns. The queries used in Günther's study were designed to identify nominal ellipses in the British National Corpus (Günther, 2003: 86ff.) and we tested similar queries adapted to the different tag sets used in our data to identify nominal ellipses in our English and German data.

We considered Günther's corpus study as a promising source on the methods for identifying nominal ellipses in texts that are tagged for part-of-speech categories. Finally, the differences with regard to the size of the BNC and GECCo did not make it possible to use similar methods for our data. Günther used a large monolingual corpus where a certain query for part-of-speech patterns can provide numerous examples that are relevant for the description of a pattern of interest even if many, if not most, instances are missed. In her study, several difficulties of the automatic search for elliptical noun phrases were circumvented by the exclusion of many items from the search if the results led to many irrelevant examples. She only looked for elliptical noun phrases that have the pattern of *'the'*,

'a', 'and', or 'no' as a determiner, followed directly by an adjective and then by another part of speech from a selection of word classes which are not nouns. Several word classes that theoretically can occur in an elliptical noun phrase as well as longer or different patterns than the chosen two elements in a noun phrase remnant (a specific determiner and an adjective) were excluded. Additionally, Günther decided to restrict her analysis to the 1,000 most frequent adjectives in the patterns that the query detected and therefore excluded countless other potential candidates for nominal ellipsis such as ellipses after numerals, possessives, classifier nouns, quantifiers and after pre-modified adjectives due to practical reasons. While Günther's theoretical part of her book on the elliptical noun phrase in English covers several types of nominal ellipsis, the corpus analysis concentrates only on one possible pattern which is not necessarily the most relevant pattern, but one that yields the best results. Using this method makes sense given the size of the BNC and the aim to find examples that illustrate a point. This methodology cannot be applied to our corpus as it is much smaller than the BNC and it risks excluding the majority of relevant cases.

Additionally, one would think that the query used in the BNC had low recall, but a high precision, but it still yielded numerous irrelevant examples. Therefore it resulted in relatively low precision *and* low recall. We made the same observation when trying out similar queries in the GECCo corpus. They have the advantage of rapid query formulation, but many hits have to be sorted out manually in a labour-intensive process. Additionally, every example has to be analysed in its context to distinguish

nominal ellipsis from other phenomena. In our case, as we are interested in nominal ellipsis as a cohesive device, we additionally have to check each context for possible antecedents which can occur in a different sentence. Moreover, in Günther's study as well as in this study, it turned out to be even more difficult to spot nominal ellipses automatically across a wide variety of registers, particularly in spoken data, so that automatic query results may overrepresent examples found in written texts, as they can be identified more precisely with automatic queries.

One important aspect that finally led to the decision to use manual annotation is the observation that part-of-speech tags generally tend to be incorrect in ellipsis environments and in other contexts that involve deficient or non-standard syntax (such as spontaneous spoken language) as well as in certain registers with numerous spelling mistakes and grammatical errors (e.g. our data from online discussion forums). This implies that several relevant ellipsis cases can never be found with automatic methods if wrongly tagged elliptical structure cannot be anticipated. Many adjectives in incomplete noun phrases, for instance, are wrongly tagged as nouns both in the BNC and in GECCo. This type of elliptical phrases cannot be identified with a part-of-speech-based query for *article + adjective not followed by a noun* (e.g. `[pos='art'] [pos='adja'] [pos!='nn']` in GECCo²⁰⁵) if the adjective in the

²⁰⁵ The English tag set in GECCo is mainly based on the Penn Treebank Tagging Guidelines (<http://www.clips.ua.ac.be/pages/mbsp-tags>, last checked 08/02/2016) and the German tag set is based on the Stuttgart-Tübingen-TagSet STTS (<http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/GermanTagsets.html>, last checked 08/02/2016). One newer version of GECCo is entirely annotated with the universal POS tagset (Petrov

ellipsis remnant is incorrectly tagged as a noun. Other words that are typically wrongly tagged in elliptical noun phrases are polysemous words and homographs that can belong to different word classes, which is a relatively frequent occurrence in English. Such words normally do not pose a problem for human readers who can disambiguate them in their contexts even if several readings are theoretically possible, but taggers typically select wrong tags for such cases in incomplete noun phrases as in (8:2-3) from the GECCo corpus.

(8:2) *And he turned from the gold lady and would have taken the silver [].* (EO_FICTION_002)

(8:3) *The second [] was bright with silver.* (EO_FICTION_002)

In (8:2), *'silver'* is used as an adjective with the meaning 'having the colour of silver'. The phrase cannot be found with corpus queries for incomplete noun phrases as it is tagged as a noun. In (8:3), the nominal ellipsis after the numeral cannot be found by using a corpus query for incomplete noun phrases as *'second'* is also wrongly tagged as a noun. Moreover, a query for cardinal numbers that are not followed by nouns gives us false positives, for instance in the example (8:4) where *'second'* is a noun, but was wrongly tagged as a numeral instead. This is a particular problem in English which has many homographs and uses conversion or

et al., 2012) aimed at easing comparability of the categories under analysis across languages at the cost of granularity.

zero derivation frequently.²⁰⁶

(8:4) *Hydrogen can form stronger bonds with oxygen yet still be mobile. Thus in water, H₂O, the hydrogen atoms exchange between different oxygen atoms billions of times per second.*
(EO_POPSCI_009)

The ‘X-is-the-new-Y’ construction (e.g. ‘*X is the new black*’, ‘*X is the new normal*’) is another example of false positives in corpus queries for phrases without nouns. De-adjectival nouns tend to be tagged as adjectives in sentences such as (8:5).

(8:5) *Was this going to be her new normal?*²⁰⁷

Another example for frequent tagging errors is ‘s’ which is often tagged as a possessive marker in English, even if it is used as a shortened form for ‘has’ or ‘is’. Thus, nominal ellipses after possessive markers cannot easily be spotted automatically as several relevant cases will not be found while the query results in a long list including about 90 percent irrelevant cases. A query could find cases such as (8:6) but not (8:7) as ‘*Boeing*’ was not recognised by the tagger as a proper noun but classified as a preposition.

²⁰⁶ ‘*Exchange*’ is another word in this example that has been tagged wrongly. It is used as a verb here, but was tagged as a noun.

²⁰⁷ Example from the novel *Purity* (2015) by Jonathan Franzen, p.531, <https://books.google.de/books?id=cT94BwAAQBAJ&pg=P531> [last checked 05/03/2016]

(8:6) *Our tax rates are comparable to Germany's.* (EO_ESSAY_013)

(tagged as: ... comparable_jj to_ii Germany_np1 ' _yi s_gp)²⁰⁸

(8:7) *Their total defense business exceeds Boeing's.*

(EO_SPEECH_009) (...Boeing_ii ' _yi s_gp)

Additionally, many wrongly tagged potential triggers of lexical verb ellipsis in English are verb contractions such as *'isn't'*, *'aren't'*, *'wasn't'*, *'don't'*, *'didn't'*, *'shouldn't'*, *'wouldn't'*, *'wanna'* or *'gotta'* or unconventional spelling variants, for instance, when the pronoun *'I'* is not written as a capital letter as it often happens in English FORUM texts. This makes it difficult to find verbal ellipses that often appear in connection with contracted verbs or in sequences containing a personal pronoun. For German, we could equally list some potential common ellipsis triggers that are tagged wrongly in many cases.

If we want to consider and anticipate such tagging errors in GECCo in CQP queries, such queries will soon become rather complex and lengthy as well as considerably unfit for application in other corpora. In the context of this study, I tried to correct part-of-speech tags that turned out to be incorrect, but I came to the conclusion that there is no possibility to manually or automatically correct certain part-of-speech tags in GECCo

²⁰⁸ The English tagset in GECCo is slightly more fine-grained than the above mentioned Penn Treebank Tagging Guidelines and it has some slightly different labels for certain categories, e.g. 'gp' for possessive markers instead of Penn's 'POS' or 'ii' only for prepositions instead of 'IN' for prepositions or subordinating conjunctions.

that regularly or occasionally proved to be difficult for the tagger that was used. This would unintentionally affect other existing annotation levels at the same time. We can only work with the corpus now as it is and cannot modify the basic annotation layers.

I have provided a rather detailed insight into several initially promising strategies for identifying potential ellipsis contexts automatically in corpus data and demonstrated why (semi-)automatic methods and corpus queries cannot lead to good results in a study on cohesive ellipsis. From all ellipsis types, nominal ellipses as omissions where one particular element is left out within noun phrases should be easier to query than other ellipsis types and it already revealed many obstacles. It would be even more difficult to identify verbal and clausal ellipses automatically as they involve a larger variety of possible underlying patterns than nominal ellipses. As the examples of ellipsis in GECCo given in Chapter 5 demonstrate, cohesive ellipses occur in a large variety of remnant structures; in this study I am also interested in the question of how many ellipses there are in which text type and in the German and English data in general. It is highly desirable to achieve both high precision and recall in this study in the light of the relatively low number of cohesive ellipses that has been suggested by previous corpus-based studies on ellipses. My aim was to create reliable annotations and to find all relevant examples in the corpus. I did not want to miss any instances of the patterns of interest by queries or to obtain query results with too many false positives that I would have to sort out by hand. Not all possible ellipsis contexts and ellipsis-antecedent relations can

be anticipated, ranging from omissions of single words to omissions of larger parts of clauses linked to antecedents that can occur in the same sentence or elsewhere in the preceding or following context and that even may involve certain grammatical mismatches. Queries for specific potential ellipsis triggers usually lead to a majority of irrelevant cases that have to be excluded manually. The analysis of the context of each query result for potential antecedents of ellipses would be time-consuming and complicated with automatic methods.

I therefore opted for a completely manual annotation for ellipses and fragments, similarly to what is currently done in the annotation process of lexical cohesion in GECCo where automatic pre-annotations proved to be more misleading than helpful. Although the manual annotation procedures for ellipsis took some time, they have been found less time-consuming than the manual annotation of lexical cohesion, a cohesive device that occurs much more frequently and is equally complex and not entirely straightforward to define. Manual annotation of a complex phenomenon such as cohesive ellipsis is much more accurate than automatic methods that would inevitably have a lower precision and / or recall.

Nevertheless, the annotations that were created and the extracted patterns can serve as a basis for similar annotations in a larger corpus in the future. The most frequent and most relevant grammatical patterns of the identified ellipsis contexts in each register can help to improve semi-automatic annotation methods for specific ellipsis types in a particular language.

9. Analysis of corpus data and results

9.1 Overview, extracted results and computational tools

This chapter will document the results of the quantitative analysis of the corpus data. The major motivation of this thesis is to better understand the role of ellipsis as a cohesive device and how close German and English are in their linguistic systems of cohesion with regard to the use of ellipses. The frequencies of other types of ellipses that do not primarily function as cohesive devices as well as other fragments will only be provided for comparative purposes. We will first describe and interpret the findings from a comparison of the English and German non-translated data. Then we will give an overview on the comparison of original texts and translated texts. Finally, we will have a closer look at the different registers and the differences between written and spoken mode. We will provide frequency data and generate summary statistics, visualise these graphically with different types of plots, interpret the data and test various comparative and relationship hypotheses.²⁰⁹ The GECCo corpus allows us to examine

²⁰⁹ In the context of the GECCo project, various hypotheses have been developed and tested on the proportion of cohesive devices depending on the variables register, mode, translation status and language. Some aspects of the distribution of cohesive devices in the original texts of the GECCo corpus have been addressed in Lapshinova-Koltunski and Kunz (2015). Coreference and conjunctions as highly frequent cohesive devices were analysed in this publication and compared with each other and the use of substitution. Kunz et al. (submitted) focus on some shallow features as indicators of English-German contrasts in lexical cohesion and Kunz et al. (forthcoming a) address the interplay of several types of cohesive devices, including ellipses, which are considerably less frequent than other types of cohesive devices. In these studies, various types of quantitative analyses such as descriptive data analysis, correspondence analysis and classification with support vector machines were used. Not all types of quantitative analysis that have been used for the analysis of several cohesive devices at the same time or specific, frequent cohesive devices can be applied when focussing on ellipsis as a specific cohesive device.

the overall differences between original English and German texts on the basis of the largest dataset within the corpus sections. For the identification of the characteristics of translations versus original texts, written versus spoken mode and for the analysis of register-internal and cross-registerial variation, we have to work with smaller datasets within the respective subcorpora. We extracted all passages containing cohesive or non-cohesive nominal ellipses and verbal/clausal ellipses²¹⁰ as well as their translations from the corpus.

Each example has been examined in detail to determine the specific structure of each ellipsis on a more-fine grained level. Additionally, we collected information on the cases where a number mismatch occurred in the examples and on the structures that were used in the translations of these passages. We came to the conclusion that most ellipsis-antecedent-relations do not stretch beyond two sentences. Most cases of ellipses in both English and German are relatively local phenomena – whether they are used as cohesive devices or not.

Annex 1 lists the frequencies of each ellipsis type per subcorpus, register and text in GECCo. The tables include the number of tokens in the respective section of the corpus, the absolute frequencies of cohesive

²¹⁰To subsume verbal and clausal ellipsis under one heading here makes sense as the frequencies per texts are low. As explained in the previous chapters, verbal and clausal ellipses have been annotated separately. In the analysis, however, I chose to treat them under one category as they are conceptually relatively similar and the annotation process proved that they are not very frequent when considered separately. Mixed cases of co-occurrence of nominal and verbal/clausal ellipsis were also annotated separately, but they are so rare in our data that I subsumed them under nominal and verbal/clausal ellipses. Several non-cohesive verbal und clausal ellipses also establish cross-clausal relations and some might argue that they therefore establish textual relations, but they have not been listed under cohesive ellipses if they are merely the result of coordination.

ellipses and their subtypes, of non-cohesive ellipses and other fragments. Additionally, the raw frequency values were scaled to a frequency per 1,000 tokens for individual texts and to a frequency per 10,000 tokens for the registers and subcorpora to obtain normalised frequencies which are also provided in the tables in Annex 1.²¹¹

In the current analysis, I do not focus on more fine-grained subcategories than nominal and verbal/clausal ellipsis as we do not have high numbers of occurrences for the more fine-grained subtypes. This means I will not compare details of nominal ellipsis after numerals vs. nominal ellipsis after adjectives, lexical verb ellipsis vs. modal verb ellipsis, sluicing as a subtype of clausal ellipsis vs. clausal ellipsis in question-answer pairs or in other rejoinder sequences, etc. For practical reasons, a quantitative analysis of more fine-grained labels indicating specific subtypes and ‘sub-subtypes’ will not be feasible in this study as there are numerous types of structures for which we find only a few examples in the data and I do not wish to imply that findings based on a low number of occurrences should be used to make generalised statements

²¹¹ I also tested if it would make sense to express the numbers of nominal and verbal/clausal ellipses per numbers of noun phrases, verb phrases or clauses as these are the contexts in which ellipses of the different types are possible. Finally, I opted only for normalised scores per tokens to make it easier to compare our different categories and to provide more meaningful results as the texts for the corpus had been selected with regard to their similarity in text length and not with regard to their density of NPs, VPs or their number of clauses. It is not straightforward to count NPs, VPs and clauses in the data as many texts have quite complex nested structures. It is possible to query elements of NPs, VPs and clauses in the GECCo corpus, but NPs that have been annotated also include pronouns that cannot contain ellipses of head nouns. Often NPs that are embedded in other structures have not been marked as separate NPs in the annotation. Different types of verbal elements have been annotated separately (finite elements of VPs and other elements of VPs, infinitives and gerunds). Therefore, a query for VPs or for one of these structures does not provide the overall number of verb phrases, but only of its elements. Moreover, phrasal structures in the written data have been annotated with a different method than in the spoken data.

about the English and the German language as a whole. Nevertheless, more detailed information on the respective subtype of each example is given in the annotation scheme to describe the phenomena that fall under the respective headings. It is possible to select only certain specific ellipsis patterns from the results by either querying subtypes or by taking certain types of examples from the lists of extracted ellipses that were generated and supplemented with more fine-grained information on each case. It is also possible to combine a query for the annotated ellipsis category in combination with other query constraints with regard to other corpus annotation levels (e.g. tokens, lemmas, morpho-syntactic features, parts of speech, phrase chunks and their grammatical functions or sentence boundaries) and to see how specific subtypes of the main ellipsis categories are distributed in the data. Numerals and quantifiers, for instance, often occur in parallel structures indicating contrast or showing comparison and therefore are frequent in remnants of nominal ellipsis. Words that are used as numerals and quantifiers are not as diverse as open-class descriptive adjectives; they tend to occur frequently in different registers and are rarely followed by nominal substitutes. Additionally, numerals can occur in sequences across sentences (e.g. ‘the first’, ‘the second’, ‘the third’ etc.) so that one antecedent noun can be followed by several cohesive elliptical phrases. The use of the subtypes of nominal ellipsis therefore also depends on the field of discourse and topic continuity. Although some would probably argue that cases that we group together differ with regard to their internal and remnant structures, we

think that it has more advantages to group them together in overarching categories.

Our quantitative analysis was carried out with a combination of tools. We mainly used *IBM SPSS Statistics 21*²¹² and *MATLAB R2015/2016a*²¹³ to analyse and visualise the data. In corpus linguistics, very often *R* is used as an open-source tool environment for statistical computing and graphics (Gries, 2009, 2013). I decided to use SPSS and MATLAB as they are, like *R*, well-documented standard tools for processing, visualising and analysing large amounts of data. These programmes can be used to extract patterns, to identify trends and relationships between variables and to test hypotheses. MATLAB is both a programming language and a software environment. The most typical users of MATLAB are engineers, but the MATLAB working environment turned out to be well-suited for most of our purposes, particularly for representing the data visually. It offers a set of tools for importing data, e.g. from Excel Spreadsheets, and for managing variables in the workspace. It is possible to interactively preview, select, and pre-process the data before importing them. An extensive set of built-in math functions and interactive plotting functions supports the statistical analysis of data and their visualization. An additional Statistics and Machine Learning Toolbox includes a greater variety of statistical functions than the standard version. SPSS on the other hand, a widely used statistical analysis software in social sciences and

²¹² <https://www-01.ibm.com/software/de/analytics/spss/products/statistics/> [last checked 28/04/2016]

²¹³ <http://de.mathworks.com/products/matlab/> [last checked 28/04/2016]

business research, can quickly perform all types of statistical calculations requiring less programming effort than MATLAB. We used it for some calculations of basic statistics and to check some results of the MATLAB calculations. The following sections of this chapter will present and examine data that are also given in Annex 1 and focus the findings on the research questions and hypotheses discussed in Chapter 1.3.

9.2 Cohesive ellipses in English and German

This section examines the frequencies of cohesive ellipses in the English and German datasets of original texts in GECCo.

Table 5 lists the absolute and normalised frequencies of cohesive ellipses and its subtypes in EO (English originals) and GO (German originals).

Sub-corpus	tokens	all coh. ell. (abs.)	all coh. ell. (per 10,000 tokens)	coh. nom. ell. (abs.)	coh. nom. ell. (per 10,000 tokens)	coh. verbal / clausal ell. (abs.)	coh. verbal / clausal ell. (per 10,000 tokens)
EO	408,016	397	9.73	211	5.17	186	4.56
GO	414,027	397	9.59	207	5.00	190	4.59

Table 5: Absolute and normalised frequencies of all cohesive ellipses, cohesive nominal ellipses and cohesive verbal/clausal ellipses in English and German originals in GECCo²¹⁴

As expected, the absolute frequencies of cohesive ellipses are not very high in either language, but surprisingly, the total numbers of all types of cohesive ellipses in English and German originals are exactly identical with 397 occurrences in each corpus sections that contain ca. 400.000 tokens each. There are 4 nominal ellipses more and 4 verbal/clausal ellipses less in English compared to German originals. If we compare the normalised frequencies, we can see that English originals have only marginally more cohesive ellipses than German originals. These subcorpora contain about 9.7 (English) and 9.6 (German) cohesive ellipses

²¹⁴ abs. = absolute, coh. = cohesive, ell. = ellipses, nom. = nominal

per 10,000 tokens. The number of cohesive nominal ellipses is higher than that of cohesive verbal and clausal ellipses in both languages. The English and German corpus sections contain about 5 cohesive nominal ellipses per 10,000 tokens and about 4.6 verbal and clausal ellipses per 10,000 tokens.

The stem plot in Figure 16 shows how cohesive ellipses are distributed across the individual texts in English and German. For convenience and in relation to the length of the texts, the frequencies have been normalised to occurrences per 1,000 tokens for each text in this figure (and not per 10,000 tokens as was done in the description of the subcorpora). The texts in this figure have been sorted by the frequency of cohesive ellipses.

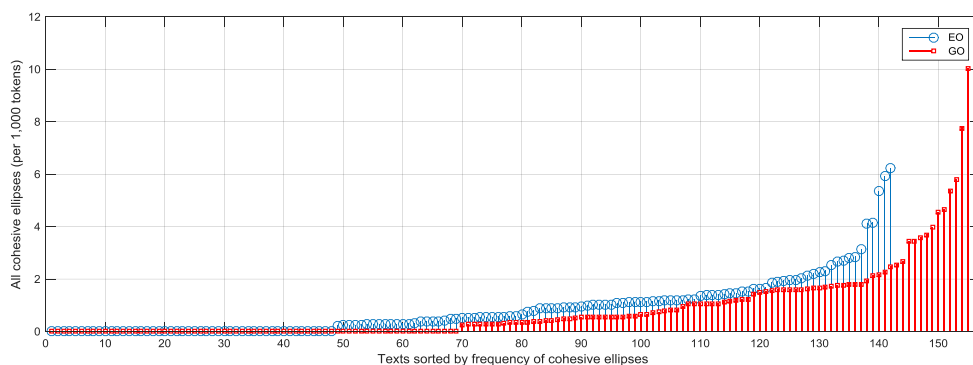


Figure 16: Number of cohesive ellipses per text in EO and GO (normalised per 1,000 tokens, texts sorted by frequency of cohesive ellipses)

Figure 16 demonstrates that the distribution of cohesive ellipses is relatively similar in English and German. The values are between 0 and 2 (per 1,000 tokens) for most texts (126 out of 142 text in English and 138 texts out of 155 in German, which amounts to 89% of the texts in each language – another aspect where the English and German results resemble

each other closely). Both English and German have about 10% of the texts with values between 2 and 6 and about 1% of the texts with values higher than 6. Among the German texts, there are 2 texts with the highest values of all texts (ca. 8 and 10). Moreover, in the German data we find a higher proportion of texts with no cohesive ellipses at all (69 out of 155 (45%) compared to 48 out of 142 (34%) in English).

The average or mean values for EO and GO that were calculated on the basis of these data are almost exactly the same (0.873 for English and 0.872 for German), but the median, i.e. the value that separates the lower 50% of the distribution from the upper 50%, is higher and also closer to the mean value in English (median = 0.53 in EO vs. 0.34 in GO). The stem plot displayed in Figure 16 seems to suggest that the German data have a higher variability than the English data. The calculated standard deviation (σ) as an indicator of dispersion of the data confirms this observation. The standard deviation is the square root of the variance (the average squared difference of the scores in a distribution from the mean). $\sigma = 1.14$ for English and 1.46 for German.²¹⁵

Our data are not normally distributed. If a variable has a normal distribution, 68.26% of the values fall within one standard deviation from the mean, 95.44% within two standard deviations, and 99.74% within three standard deviations from the mean. However, 89% of the values for the

²¹⁵ The formula that is used in MATLAB is given under <http://de.mathworks.com/help/matlab/ref/std.html> [last checked 10/05/2016]. The standard deviation is a good measure of spread relative to the mean if the data are not skewed and heavily affected by outliers. It is somehow debatable whether this is the case here with many zero values and certain outliers. We should also keep in mind that there are fewer texts in the English corpus section than in the German one which could have contributed to the lower number of outliers in English.

English data fall within one standard deviation from the mean (all values from 0 to 2.01). 91% of the values for the German data fall within one standard deviation from the mean (all values from 0 to 2.33). In EO, 96.5% fall within two standard deviations from the mean (values from 0 to 3.15), in GO 95.5% (0 to 3.79). Values lying more than two standard deviations either below or above the mean can be considered as outliers, which gives us 3.5 to 4.5% of minor and extreme outliers for our data. We even have about 2 to 3% of extreme outliers with values higher than three standard deviations from the mean and lie above these thresholds of 4.29 in EO and 5.25 in GO.

The corpus texts vary in length and it is useful to examine whether the text length correlates with the number of cohesive ellipses and also whether many of the zero values can be explained by the brevity of the respective texts. As discussed in Chapter 7.2, there may be a risk that in several sample texts the phenomenon of cohesive ellipsis simply does not occur because the texts are too short. The scatter plots in Figure 17 visualise the relationship between text length and frequency of cohesive ellipses for EO and GO.

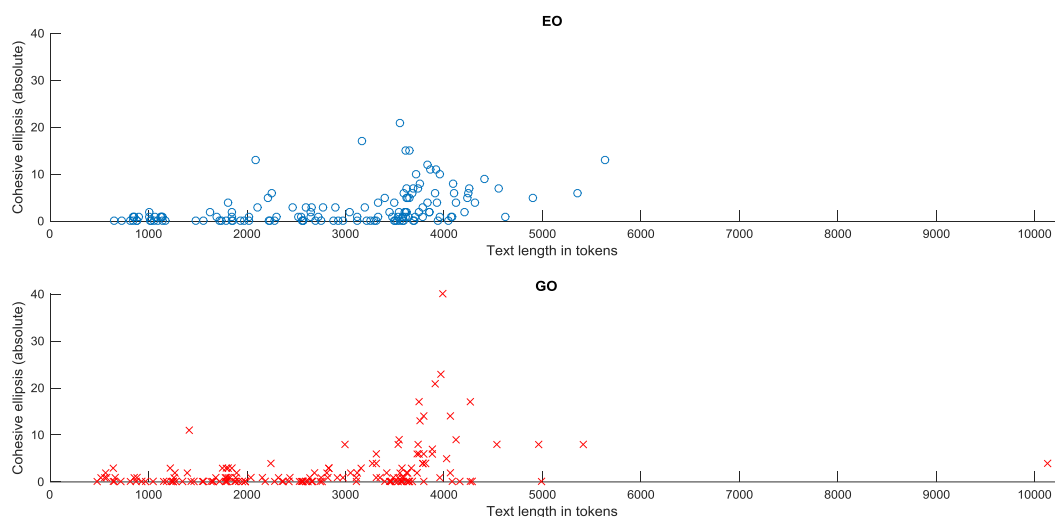


Figure 17: Scatter plots for text length vs. absolute number of cohesive ellipses in EO and GO

Figure 17 shows that it is not only the short texts that have low absolute ellipsis values and it is not only the long texts that have high absolute values either. In both languages, we can see that the texts with zero values vary greatly in length and that many of the shortest texts do have low values, but not necessarily zero values. However, texts longer than 3,000 tokens here generally have higher values; only few texts which are longer than 4,000 tokens have zero values. To determine quantitatively the extent to which the variables text length and number of cohesive ellipses are correlated, we calculate Pearson's correlation coefficient (r) as a measure of their linear dependence. The values of Pearson's correlation coefficient always lie between -1 and $+1$. If one variable increases when the second also increases, then there is a positive correlation. In that case the correlation coefficient will be closer to 1 . If $r = +1$, then there is a perfect

positive correlation, if $r = 0$, there is no linear correlation and if $r = -1$, then there is a perfect negative correlation.²¹⁶ The p-value can be used to test the hypothesis that there is no relationship between the observed phenomena. The p-value is the probability to find the current result if the correlation coefficient is actually zero. If it is smaller than the significance level (by convention: 0.05),²¹⁷ then the corresponding correlation is considered statistically significant. For EO, $r = 0.43$ (with $p = 8.09 \times 10^{-8}$) and for GO, $r = 0.35$ (with $p = 7.25 \times 10^{-6}$). The values indicate a low to moderate degree of positive correlation between text length and number of cohesive ellipses. The correlation coefficient is slightly higher in German

²¹⁶ It is easy to misinterpret a correlation coefficient. Therefore I wish to make some remarks to clarify some concerns. A non-zero correlation coefficient indicates that the numbers are related, but only if the coefficient is either 1 or -1, there may be other influences that have an impact on the relationship between the variables. The closer a correlation coefficient is to zero the greater the uncertainty. A zero correlation coefficient can be an indicator of a true non-correlation, but a zero correlation can also be obtained if positive and negative correlations of subgroups overlap (cf. Grawe, 1991: 96f.). Additionally, it should be kept in mind that a zero value of the correlation coefficient does not necessarily mean that there is no relation at all. It only specifies that there is no linear relation. Therefore a zero correlation coefficient may hide a different type of relation and does not necessarily imply that two variables are genuinely independent. A high correlation coefficient on the other hand does not necessarily imply that there is a large effect and that a change in one variable will produce a large change in the other variable. Correlation coefficients can be affected by extreme outliers. If we exclude outliers, the correlation drops or rises, but by doing that we would manipulate the corpus data in a certain way. Like p-values, correlation coefficients are subject to sampling variation. With larger datasets, we should set the threshold for the r-value higher to see a strong correlation in the data.

²¹⁷ In empirical studies, the p-value for determining statistical significance is arbitrarily set in advance. It can take any value between 0 and 1. Due to the way it is calculated, smaller p-values are obtained with larger sample sizes. With bigger datasets, one should normally set a lower threshold for the p-values to interpret something as significant. Therefore I also considered determining a p-value threshold of 0.01 or 0.001 for analyses of larger sections of the corpus and 0.05 for smaller sections, but I decided to set it at 0.05 as per convention in corpus-linguistic studies. The interested reader is referred to Ziliak and McCloskey (2008) who raise some important caveats with regard to significance testing in their book *The Cult of Statistical Significance*.

than in English and the p-values for both datasets suggest that the results are statistically significant.

Let us now have a closer look at the ellipsis subtypes nominal ellipses and verbal/clausal ellipses. Figure 18 shows how cohesive nominal ellipses are distributed among the texts in EO and GO. The thin dotted lines in the bar plots indicate register boundaries and the thicker line indicates the boundary between written and spoken mode²¹⁸. We will zoom in on the data with regard to a comparison of the registers and between written and spoken mode in Chapters 10.4 and 10.5.

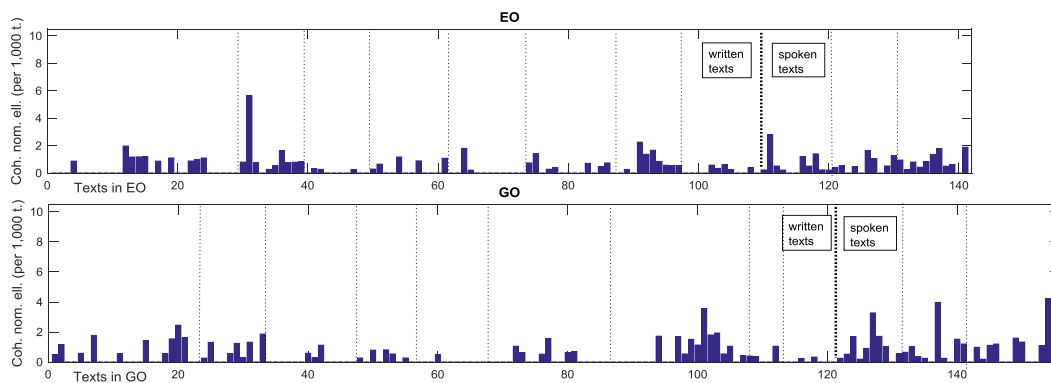


Figure 18: Cohesive nominal ellipses in EO and GO

In Figure 18, where we look only at nominal ellipses at one subtype of cohesive ellipses, we observe relatively low values in the individual texts. In Table 5, the absolute and normalised frequencies for cohesive nominal ellipses were provided. In English originals, there were 211 cases (5.17 per

²¹⁸ The texts have been sorted according to their corpus text ID in alphabetical order with a separation between written and spoken texts as in Annex 1: ESSAY, FICTION, INSTR, POPSCI, SPEECH, TOU, WEB, ACADEMIC, FORUM, INTERVIEW.

10,000 tokens²¹⁹). For English, the mean value for cohesive nominal ellipses in the individual texts is 0.48 per 1,000 tokens²²⁰, the median is 0.24; and $\sigma = 0.72$. We have 67 zero values out of 142 (47%). 88% of the texts lie within one standard deviation from the mean, 97% are within two and 99% are within three standard deviation from the mean. There is only one outstanding outlier with a frequency of 5.64 nominal ellipses. For German, the data look slightly different. We have 207 cohesive nominal ellipses and 5.00²²¹ per 10,000 tokens in the whole corpus section. The mean value for cohesive nominal ellipses in the individual texts in GO is 0.52²²² per 1,000 tokens, and the median is actually 0 because of 81 zero values out of 155 (52%). $\sigma = 0.8$. We have 86% of the text within one standard deviation from the mean. 96.7%, are within two standard deviations from the mean – we have almost the same value for those that are within three standard deviation from the mean (97.4%). That means we have more extreme outliers above this threshold (2.6%) than in English. Looking at the individual texts, the average number for nominal cohesive ellipses is marginally higher in German than in English. We clearly have more zero values, more outliers and a higher variability in German.

The bar plots in Figure 19 visualise how cohesive verbal/clausal ellipses are distributed among the texts in EO and GO.

²¹⁹ calculated in relation to the subcorpus of EO as a whole (408,016 tokens)

²²⁰ calculated in relation to the sizes of the individual texts in EO

²²¹ calculated in relation to the subcorpus of GO as a whole (414,027 tokens)

²²² calculated in relation to the sizes of the individual texts in GO

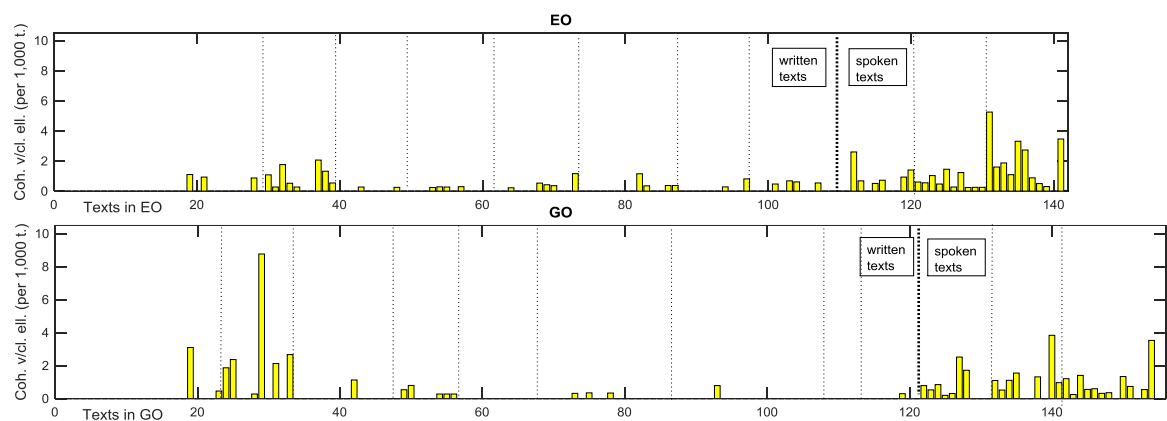


Figure 19: Cohesive verbal/clausal ellipses in EO and GO

The values for cohesive verbal/clausal ellipses are quite low in both English and German (a comparison of the registers and between written and spoken mode will be discussed in Chapters 9.4 and 9.5). The absolute and normalised frequencies for cohesive verbal/clausal ellipses in the whole corpus section of EO are 186 and 4.56 (per 10,000 tokens²²³). The mean value for this type of ellipsis in the individual texts in EO is 0.39 (per 1,000 tokens), the median is 0 and $\sigma = 0.76$. We have 84 zero values out of 142 (59%). 89% of the texts lie within one standard deviation from the mean, 96% are within two and 97% are within three standard deviation from the mean. Here, in contrast to cohesive nominal ellipses in English, much more than 50% are zero values and we have more large outliers.

In German, there are 190 ellipses of this type which corresponds to 4.59 occurrences per 10,000 tokens if we calculate it against the whole corpus section. The mean value that we obtain when looking at the frequencies for the individual texts is 0.37 (per 1,000 tokens). This value is rather similar

²²³ calculated against the whole corpus section

to the mean in the English data. The median in GO is 0 as it is in the English data and σ here is 0.97, therefore higher than in English. We have 111 zero values out of 155 (72%) in GO. This means we have many more zero values for verbal/clausal ellipses in German compared to English and also considerably more of these values than for cohesive nominal ellipses in both languages. 92% of the German texts lie within one standard deviation from the mean, 95% are within two and 98% are within three standard deviations from the mean. Here, German has fewer high outliers, but it has one extreme outlier at 8.77.

Figure 20 summarises the differences between the texts in EO and GO with regard to the distribution of the ellipsis subtypes in the form of a boxplot.²²⁴

²²⁴ Here, outliers are calculated slightly differently (i.e. not on the basis of the standard deviation σ , but on the basis of the interquartile range [IQR])<http://de.mathworks.com/help/stats/boxplot.html> [last checked 12/05/2016]

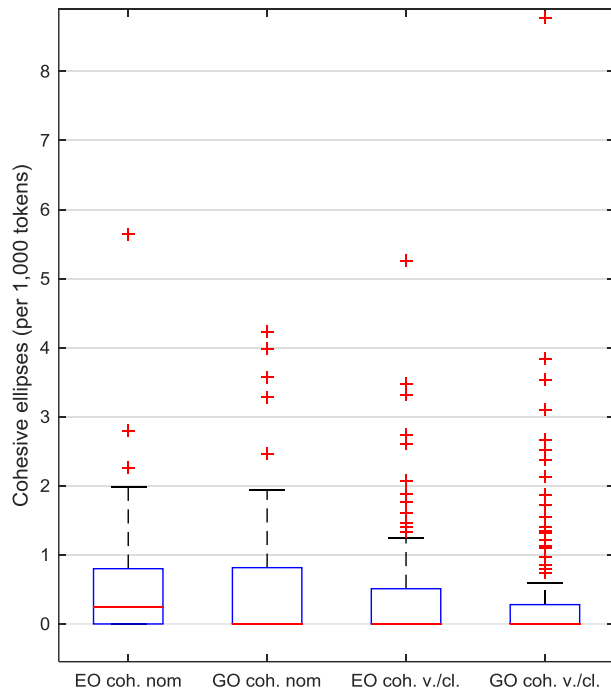


Figure 20: Cohesive nominal and verbal/clausal ellipses in texts in EO/GO

Figure 21 shows the frequencies of both types of ellipses in the texts in EO and GO in a stacked bar chart.

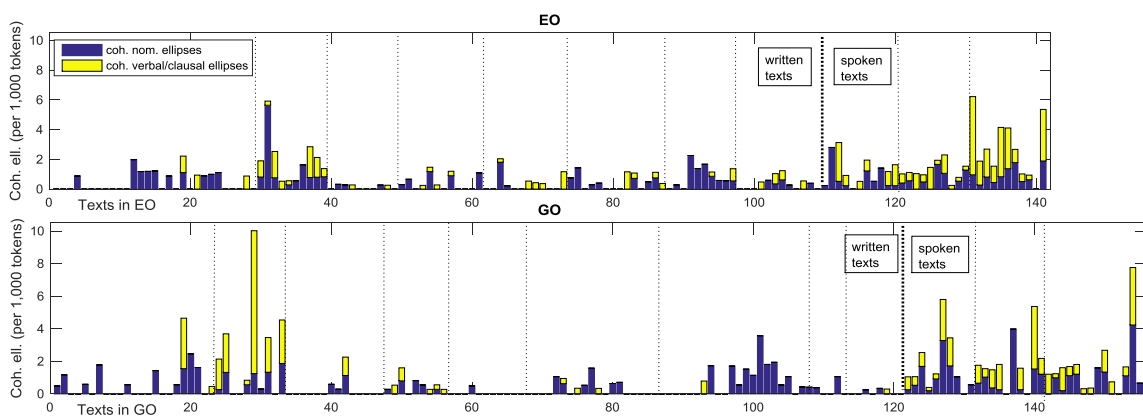


Figure 21: Cohesive nominal and verbal/clausal ellipses in EO/GO

In Figure 21, we can see that not all texts with high values for cohesive nominal ellipses also have high numbers for cohesive verbal/clausal ellipses. Although many texts with the highest frequencies of verbal/clausal ellipses also have a certain amount of nominal ellipses used as cohesive devices, it is rarely the case that a text will have only verbal/clausal ellipses, but no nominal ellipses. However, we have several texts that have only or almost only nominal ellipses. There is probably no linear relationship between these ellipsis types.

The scatter plots in Figure 22 visualise the relationship between cohesive nominal and verbal/clausal ellipses for the texts in EO and GO.

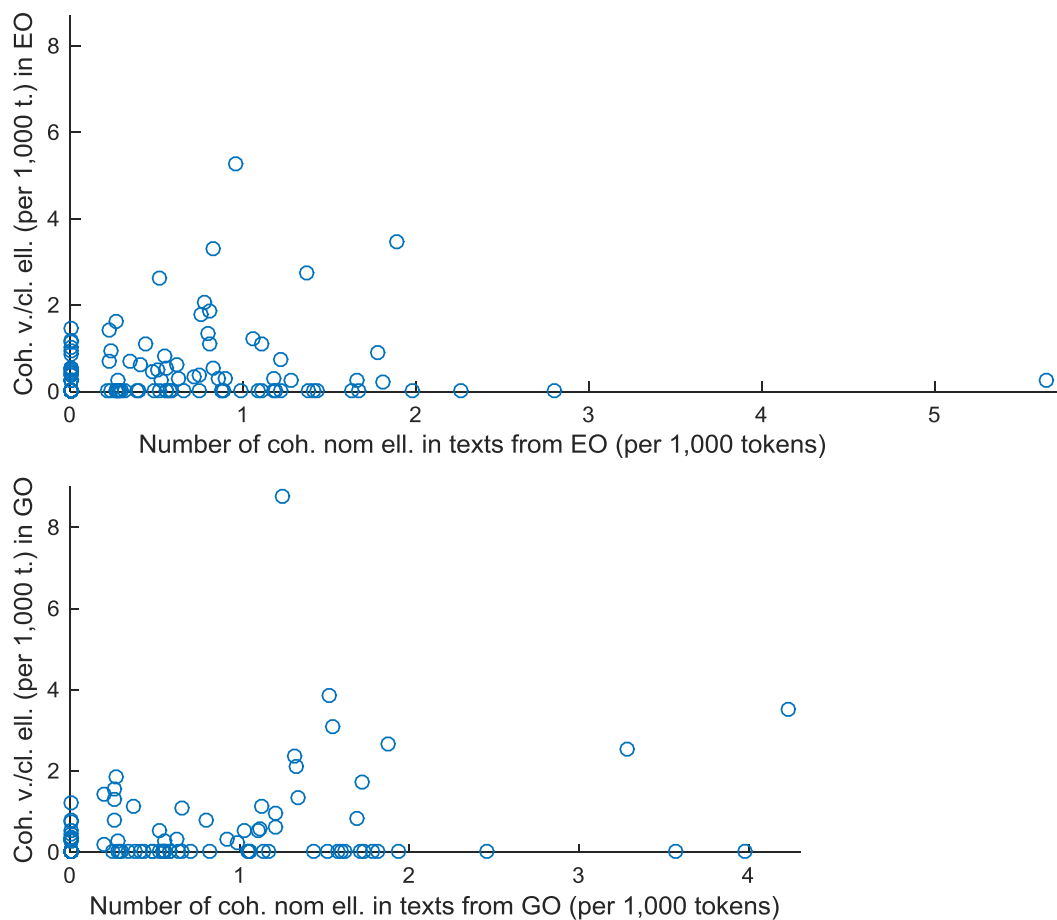


Figure 22: Scatter plots for cohesive nominal and verbal/clausal ellipses in EO/GO

Many pairs of data in these scatter plots have one zero value and may be evidence against a correlation, but for the texts with two non-zero values there could be a positive correlation. The calculation of the correlation coefficient r for EO gives us 0.16 ($p = 0.047$) and 0.36 for GO ($p = 2.88 \times 10^{-6}$). If we eliminate all data points of which at least one is a zero value as well as the data points with the highest extreme value of each variable, we still obtain an r value that is only slightly higher than zero for EO, but 0.50 for GO (the p -values for these calculations of r are higher than 0.05 and in this case we only have about 40 data points left for English and about 30 for German as most data pairs contain at least one zero value). The data seem to suggest that in English, there is no correlation between cohesive nominal and verbal/clausal ellipses, but in German there is a moderate positive correlation between these ellipsis types.

Let us now briefly turn to the frequencies of non-cohesive ellipses and other fragments in EO and GO in comparison to cohesive ellipses. Table 6 lists the absolute and normalised frequencies of all non-cohesive ellipses, its subtypes of non-cohesive nominal and non-cohesive verbal/clausal ellipses as well as the frequencies of other fragments in English and German originals in GECCo.

Sub-corpus	tokens	all non-coh. ell. (abs.)	all non-coh. ell. (per 10,000 tokens)	non-coh. nominal ell. (abs.)	non-coh. nominal ell. (per 10,000 tokens)	non-coh. verbal / clausal ell. (abs.)	non-coh. verbal / clausal ell. (per 10,000 tokens)	other fragm. (abs.)	other fragments (per 10,000 tokens)
EO	408,016	272	6.67	125	3.06	147	3.60	2168	53.14
GO	414,027	242	5.85	126	3.04	116	2.80	3181	76.83

Table 6: Absolute and normalised frequencies of all non-cohesive ellipses, non-cohesive nominal ellipses, non-cohesive verbal/clausal ellipses and other fragments in English and German originals in GECCo

If we compare Table 6 with Table 5, we see that in both languages the overall frequencies for non-cohesive ellipses and also for the subtypes of this category – nominal and verbal/clausal ellipses – are about 1/3 lower than the frequencies for the respective categories of cohesive ellipses. This means we have a relatively high proportion of non-cohesive cases among all occurrences of ellipses – a point which is often overlooked in discussions of nominal and verbal/clausal ellipses in the context of textual cohesion. The calculated means, medians and standard deviations for all non-cohesive ellipses in the individual texts per respective corpus section are almost equal in English and German (mean = 0.64 [EO]; 0.59 [GO], median = 0.45 [EO]; 0.39 [GO], σ = 0.66 [EO]; 0.71 [GO]). Particularly, non-cohesive nominal ellipses in English and German show almost no difference (125 absolute occurrences in EO and 126 in GO, which corresponds to about 3 cases per 10,000 tokens in the entire respective corpus sections). Most cases in this category are clause-internal nominal ellipses; only a few are exophoric or lexicalised. There are more non-cohesive verbal/clausal ellipses in English (147 in EO and 116 in GO,

which corresponds to 3.6 and 2.8 cases per 10,000 tokens in EO and GO). A closer look at the data revealed that it is mainly omissions of lexical verbs in gapping structures that we find among the German and English non-cohesive verbal/clausal ellipses. We also noticed that English additionally also has a relatively high proportion of lexical verb ellipses in other types of coordinated structures.

Figure 23 visualises the results provided in Table 5 and Table 6 and we can see clearly that the frequencies of the ellipses types in the English and German data are very similar.

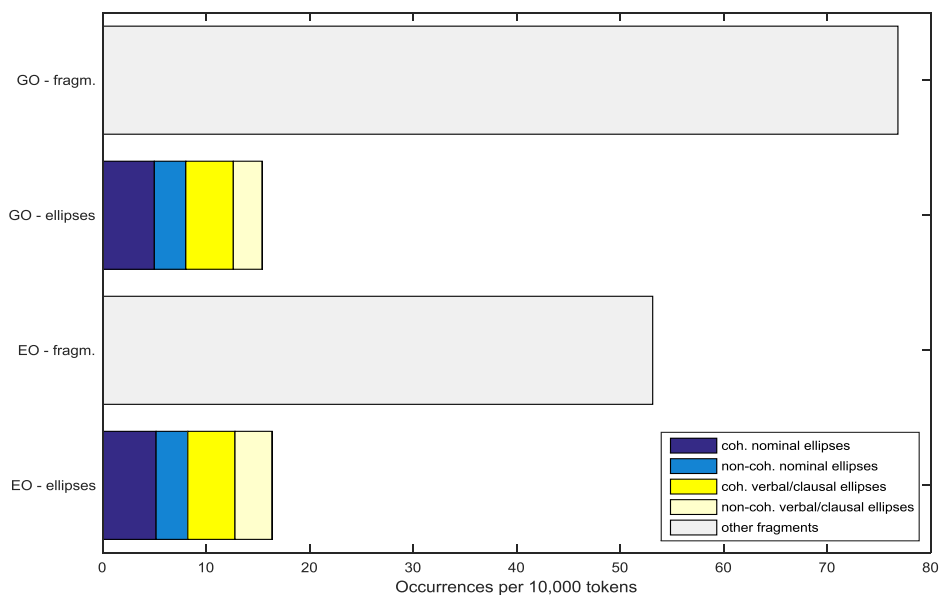


Figure 23: Cohesive ellipses, non-cohesive ellipses and other fragments in EO/GO

The scatter plot in Figure 24 shows the relationship between the frequency of cohesive ellipses and non-cohesive ellipses in the English and German texts and seems to suggest that there is a moderate positive

correlation between the number of cohesive ellipses and non-cohesive ellipses in texts in EO and GO. There are also many texts with zero values for either cohesive or non-cohesive ellipses (correlation coefficients ca. 0.35 in both languages, p-values < 0.05).

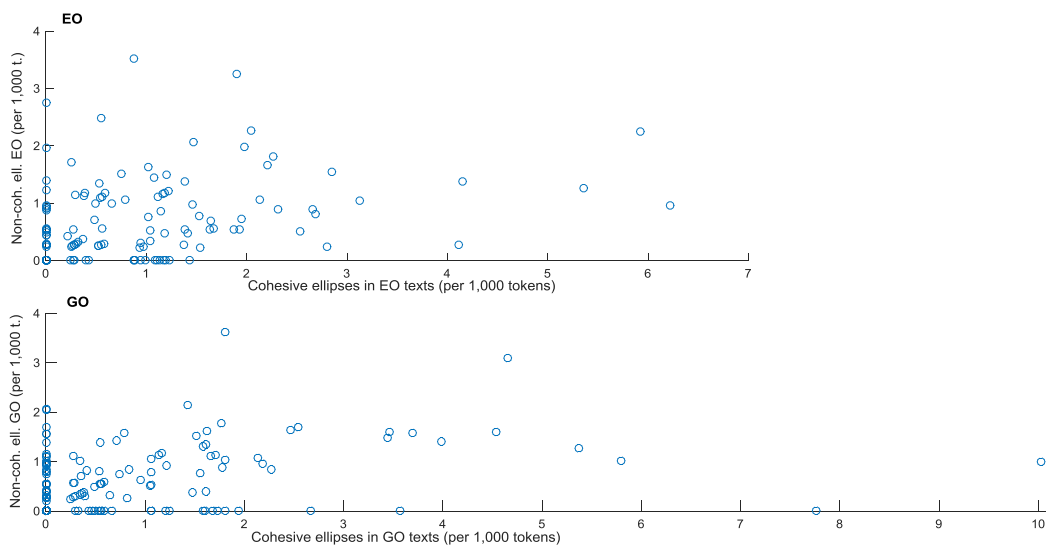


Figure 24: Scatter plots for cohesive and non-cohesive ellipses in EO and GO

We were able to show that ellipses are used in English and German with similar frequencies. There are more striking differences between the two languages if we look at the frequencies of other fragments in the texts (cf. Table 6, Figure 23, Figure 25). Fragments are considerably less frequent in English than in German (2168 vs. 3181, which corresponds to

53.14 and 76.83 occurrences per 10,000 tokens). The calculated mean values for the individual texts are 5.00 (EO) and 8.25 (GO) per 1,000 tokens. The standard deviation σ is 4.82 (EO) and 9.16 (GO). In EO, there are 9 texts with zero fragments.²²⁵ The text with the maximum value in English contains 28 fragments per 1,000 tokens. In GO, there are 3 texts with zero fragments, and the text with the maximum value in German contains 66 fragments. Here again, the data suggest a generally higher variability and more extreme outliers for the German data.

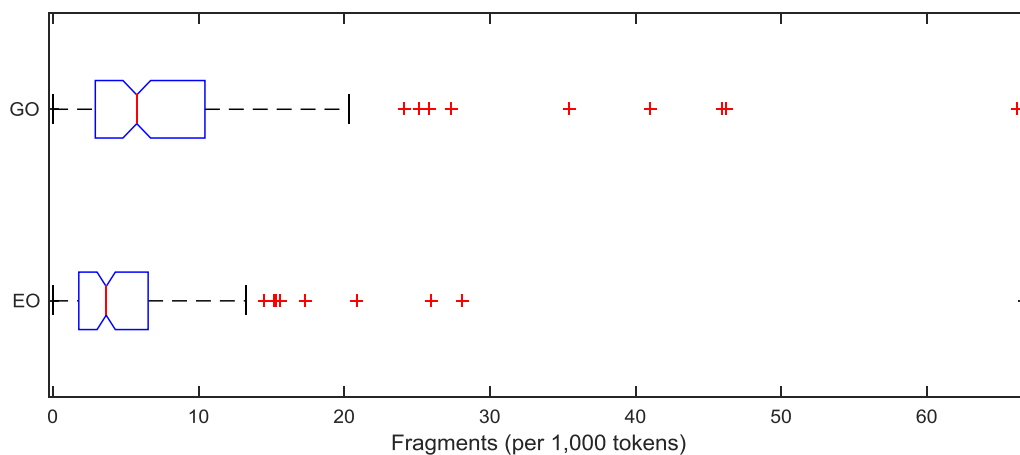


Figure 25: Notched boxplot of fragments in texts in EO and GO

The median values are 3.64 (EO) and 5.74 (GO). We therefore have a difference of about 2 between the medians of the two groups. Figure 25 is a notched boxplot where the notches represent the 95% confidence interval around the median.²²⁶ While not being a formal test, the comparison of the

²²⁵ This information is not displayed in the figures.

²²⁶ MATLAB plots the notches at the median plus or minus 1.57 times the IQR divided by the square root of the number of observations.

notches provides a rough measure of the significance of the differences between the values. If the notches of two boxes do not overlap, this offers evidence of a statistically significant difference between the medians and we can conclude with 95% confidence that the true medians do differ. Figure 25 indicates that the medians for fragments in EO and GO are significantly different at the 5% significance level as there is a non-overlap of the notch intervals.

There seems to be no particular relationship between the number of ellipses that are used as cohesive devices in the texts in EO and GO and the number of other fragments in these texts (Figure 26).

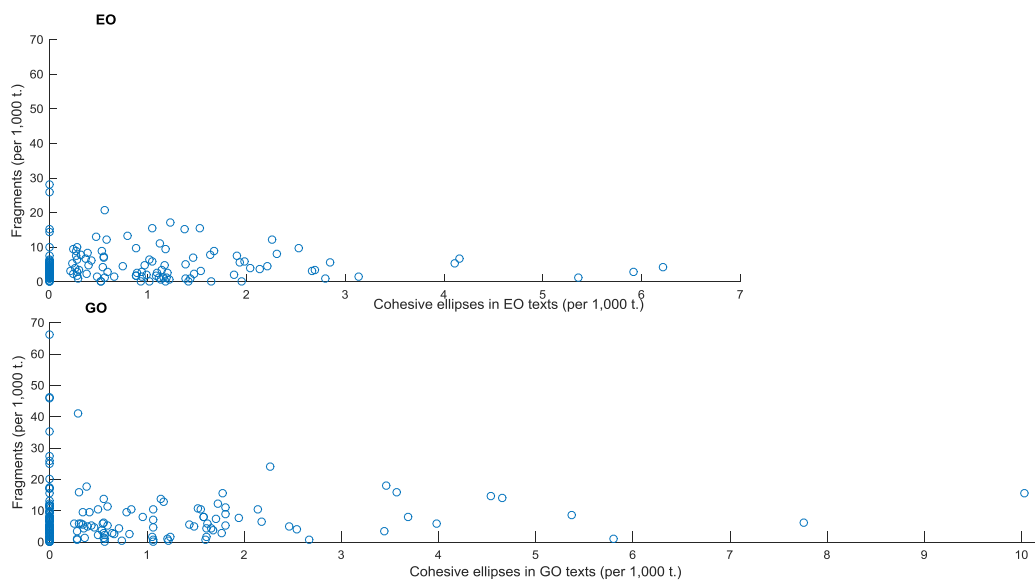


Figure 26: Scatter plots for cohesive ellipses and fragments in EO and GO

The correlation coefficients r for the data in Figure 26 are almost zero in both English and German ($p > 0.05$). Texts with either no cohesive

ellipses or with certain frequencies of cohesive ellipses have varying numbers of other fragments that do not conform to a systematic pattern. It therefore seems completely unpredictable to say how many cohesive ellipses a text will have based on the number of other fragments. Only for outliers with extremely high numbers of fragments, i.e. texts in which substantial parts are fragments, we can observe a common characteristic – they usually have no cohesive ellipses in either language.

There seem to be certain differences if we zoom in on the proportions of more fine-grained subtypes of each ellipsis category. For instance in English, nominal ellipsis remnants, particularly those of non-cohesive ellipses, often involve the same words, such as low numerals or quantifiers. However, we do not have enough occurrences to compare the languages on a statistical basis if we further subdivide nominal and verbal/clausal ellipses into more fine-grained categories than can occur in a wide variety of syntactic configurations. In conclusion, it can be said there are no substantial differences between English and German with regard to the frequencies of the main ellipsis categories we described. These structures and their main subtypes occur in both languages with similar functions. More pronounced differences between the two languages exist with respect to the frequency of other fragments.

9.3 Differences between originals and translations

This section examines the frequencies of cohesive ellipses in the corpus sections of original and translated texts. As the spoken registers do not include translations, we will analyse only the written part of the corpus in this section. It is a bi-directional parallel corpus consisting of English and German original texts (EO_written / GO_written) and their translations (GTRANS / ETRANS).

Table 7 lists the frequencies of cohesive ellipses in these corpus sections and Figure 27 visualises these results in a stacked bar chart.

Subcorpus	tokens	all coh. ell. (abs.)	all coh. ell. (per 10,000 tokens)	coh. nom. ell. (abs.)	coh. nom. ell. (per 10,000 tokens)	coh. verbal / clausal ell. (abs.)	coh. verbal / clausal ell. (per 10,000 tokens)
EO_written	286,331	188	6.57	126	4.40	62	2.17
GO_written	288,490	187	6.48	97	3.36	90	3.12
ETTRANS	322,223	148	4.59	67	2.08	81	2.51
GTRANS	284,561	184	6.47	136	4.78	48	1.69

Table 7: Absolute and normalised frequencies of all cohesive ellipses, cohesive nominal ellipses and cohesive verbal/clausal ellipses in originals (written texts) and translations in GECCo

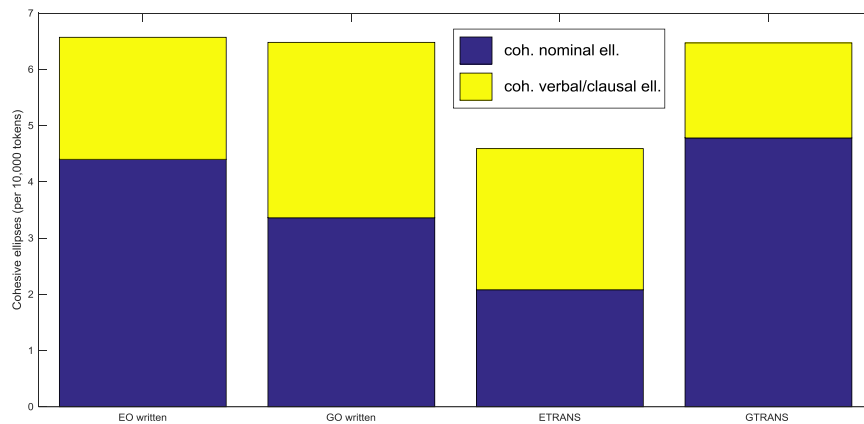


Figure 27: Frequencies of cohesive ellipses in (written) original texts and translated texts (per 10,000 tokens)

English originals have the highest number of cohesive ellipses (6.57 per 10,000 tokens), but we can see relatively similar figures for the total numbers of cohesive ellipses in the corpus sections of English and German originals as well as in the German translations (ca. 6.5 per 10,000 tokens in each corpus section). English translations have the lowest number of all cohesive ellipses (4.59 per 10,000 tokens) and the lowest number of cohesive nominal ellipses (2.08 per 10,000 tokens). German translations have the highest number of cohesive nominal ellipses, but at the same time the lowest number of cohesive verbal/clausal ellipses. In ETRANS, the number of cohesive verbal/clausal ellipsis is higher than in EO_written, which seems to reflect the influence of the source texts from GO_written with the highest frequencies of cohesive verbal/clausal ellipses. Figure 27 indicates no clear trend of generally higher or lower frequencies of cohesive ellipses in translations compared to their source texts and in

comparison to original texts of the same language. Therefore, these results only partly confirm my initial expectations to find fewer cohesive ellipses in English and German translations than in comparable original texts of the same languages and fewer cohesive ellipses in translations compared to their source texts in the other language.

These data do not represent evidence that would support the universalistic explicitation hypothesis for translations. However, the translation data suggest strong interference from the source texts with respect to ellipses as cohesive ties – which is not surprising given the close similarities in the use of ellipses in English and German described in the previous section. The translated texts show similarities to their corresponding original texts (EO_written/GTRANS and GO_written/ETRANS) as there are not substantial language-specific differences that would block or inhibit source text interference to a great extent. Figure 28 and Figure 29 display this relationship between original and translated texts on the basis of the frequencies of cohesive ellipses in the individual texts.

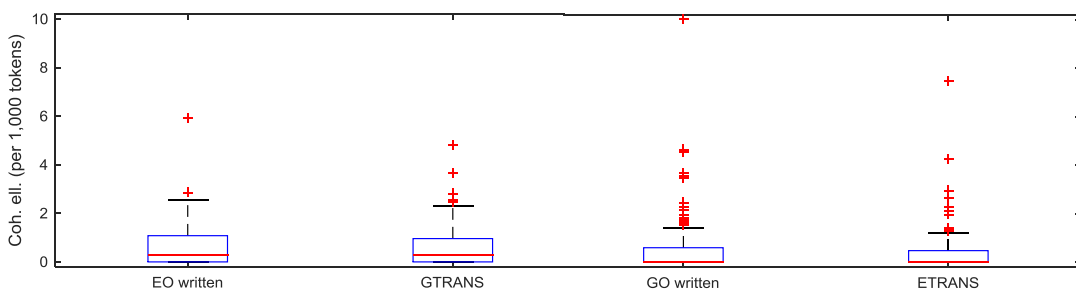


Figure 28: Cohesive ellipses in original and translated texts (per 1,000 tokens)

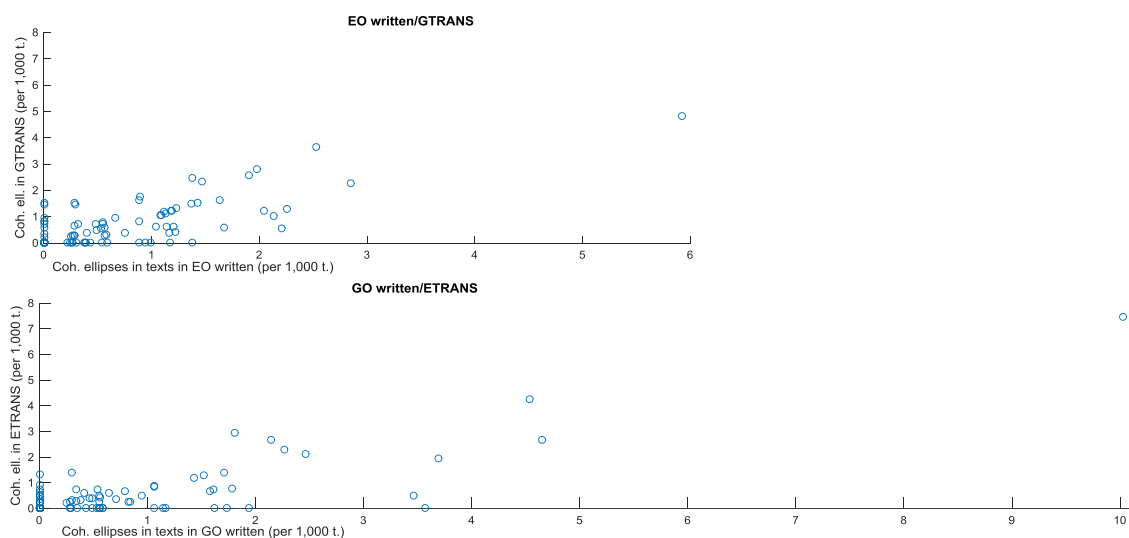


Figure 29: Frequencies of cohesive ellipses in pairs of original and translated texts (EO_written/GTRANS and GO_written/ETRANS)

The boxes in the plot for EO_written and its translation (GTRANS) as well as for GO_written and ETRANS are more similar to each other than the boxes for the two English corpus sections or the two German corpus sections in Figure 28. The linear correlation coefficient r for the data in Figure 29 is 0.78 (EO_written/GTRANS) and 0.84 (GO_written/ETRANS) with $p < 0.05$. There is a strong positive correlation between the number of cohesive ellipses in source text and their translations.

As these data do not necessarily tell us whether the ellipses in the original texts correspond to the same types of ellipses in the translations in the same text passages as in the originals, we used the sentence-aligned data and collected information on the translation strategies that were used for each ellipsis. When translating from English into German or vice versa,

a translator can often choose a similar elliptical structure in a translated text for an ellipsis from the original text. As explained in Chapter 8, we expect source language interferences with regard to the translation of ellipses used as cohesive devices. Although in certain cases, it will be difficult or impossible for translators to use an ellipsis or the same type of ellipsis as in an original text, translators will often follow the syntax of the English or German original text closely and try to omit the same material in the translation when it avoids the repetition of a textual element in close proximity to its textual antecedent. On the other hand, translators may generally tend to insert lexical material or substitutes if specific elliptical structures cannot easily be mirrored in the syntax of the translation or if they would create ambiguity, for instance due to a larger distance to the antecedent in the translation compared to the original.

By sifting through the sentence-aligned data, we found that the close relationship of English and German is apparent in our corpus as most elliptical structures from the original English and German texts were indeed translated by the use of the same type of ellipsis in the other language.²²⁷ For instance, as was shown in Table 7, English written originals in GECCo contained 126 cohesive nominal ellipses. In many cases, this type of ellipsis can also be used in the translation. Usually, translators can also use full or partial repetition of the antecedent or replace it by a substitute, a synonym, hyponym, general noun or another

²²⁷ This effect was even stronger with respect to non-cohesive ellipses where translators frequently followed the syntax of the original text more closely. Due to space limitations and the strong focus of this analysis on cohesive ellipses, I will only present the figures for the different types of cohesive ellipses in this section.

specific noun. I found that 81 ellipses of this type from EO_written were kept in the German translations imitating the syntactic structure of the original (64%). In 22 cases of ellipses in the original (ca. 17%), the translator either repeated the antecedent noun or used other means of lexical cohesion. Substitution of elided elements by semantically empty substitutes played no remarkable role in this translation direction. Neither did the replacement of elliptical structures by co-reference items such as pronouns (in both translation directions), but in some cases, certain types of modifiers in ellipsis remnants that can combine features of adjectives, determiners and pronouns are on the borderline to co-reference items in both English and German. The remaining cases in the translations from English to German were translated by the use of structures that were strongly modified during the translation process and involved entirely different types of phrasal and clausal structures. Interestingly, GTRANS contained more nominal ellipses than the original texts (136). In 55 cases, cohesive nominal ellipses that do not correspond to similar structures in the originals were inserted in the translation. Therefore, in GTRANS we have about 40% ellipses of this type where the translator omitted an element from a noun phrase from the original text in the translation. This unusually high figure of ellipsis insertions that can be regarded as implicitation strategy on the part of the translators may seem to contradict the explicitation hypothesis at first sight, but there are several reasons for this translation strategy. An ellipsis in the German translation can often be used to omit the nominal substitute 'one' for which there is no equivalent

in German and nominal ellipsis can be used in more contexts in German, particularly after adjectives (cf. Chapter 5.1.4). Moreover, some English original texts in our data are characterised by frequent lexical repetition of nouns where translators may choose to replace a repetition by an ellipsis instead.

Cohesive verbal/clausal ellipses are not as frequent as cohesive nominal ellipses in our data. There are 62 cases in EO_written of which 38 (61%) were kept in the translations, in particular the adjacency ellipses (cf. Chapter 5.3.2). Most cohesive ellipses that underwent a categorial shift in the translation process were more explicit structures involving a repetition of the lexical verb, a paraphrase of the antecedent verb phrase or clausal structure (9:1) or a general verb such as *'tun'* used as a main verb (cf. Chapter 5.2.4 and [9:2]) – often in combination with the insertion of a pronoun such as *'es'* or *'das/dies'*.

(9:1) a) *We never get it all “right” in any year. We probably never will [].* (EO_SHARE_004)

b) *Wir können nicht alles in einem Jahresverlauf richtig machen. Wir werden das vermutlich auch zukünftig nicht schaffen.*
(GTRANS_SHARE_004)

- (9:2) a) *Third, the United States must continue to press for a new round. If we do not [], no one else will [].* (EO_ESSAY_019)
- b) *Drittens müssen die Vereinigten Staaten weiterhin auf eine neue Runde drängen. Wenn wir es nicht tun, wird niemand es tun.* (GTRANS_ESSAY_019)

Only in a few cases, more implicit or shorter structures were chosen in the translation for cohesive verbal/clausal ellipses from the original text and then typically became non-clausal units or other types of fragments. Only in 10 cases, a verbal/clausal ellipsis was used in the translation that did not correspond to a similar structure in the original text.

In the translation direction from German into English, only 40 out of 97 cohesive nominal ellipses from GO_written (41%) were kept in the same syntactic environments in ETRANS. In 16 cases (16%), repetition or another means of lexical cohesion was used; in 5 cases (5%) a nominal substitute was inserted. The remaining 40% of cohesive nominal ellipses from the German original texts were translated by the use of structures that involved entirely different types of phrasal and clausal structures.

In 27 cases, a cohesive nominal ellipsis was added by the translator in the English text although it did not correspond to an ellipsis in the German original text. Some ellipses in the target texts that correspond to more explicit structures in the original text are examples of implicature or concentration strategies. ‘Implicature’ according to Delisle et al.’s (1999: 145) handbook of translation terminology is a ‘translation

procedure intended to increase the economy of the target text and achieved by not explicitly rendering elements of information from the source text in the target text when they are evident from the context'. In the case of endophoric ellipses, the implicit information is made evident by the *co-text*, the textual context surrounding an ellipsis. 'Concentration' in Delisle et al.'s terminology (ibid: 127) is a similar procedure that results in a 'decrease in the number of elements used in the target language to express the same semantic content as compared to the parallel segment in the source text. Delisle et al. (ibid.) claimed that non-English texts generally contract when translated into English. Using an ellipsis as a more implicit or more concise structure than the one used in the original text can be a useful translation strategy to simplify heavy sentence structures that are the result of the translation process. Implication and concentration are often distinguished from 'omission' as a translation procedure. An omission of text elements can be understood as a subcategory of Chesterman's 'information change' (1997) which may involve the omission of source text information deemed to be irrelevant. An omission is a deletion of information that cannot be inferred from the non-linguistic or linguistic context of the translation. Delisle et al. consider omissions as 'translation error[s] where the translator fails to render a necessary element of information from the source text in the target text' (Delisle et al., 1999: 165).

In the translations from German to English, in contrast to the other translation direction, additional ellipses that were inserted by the translator were not *mainly* cases where the translator opted for replacing lexical elements or substitutes by ellipses. These additional ellipses typically

occurred when translators used entirely different syntactical patterns than in the original. Some noun phrases, for instance, that occurred once in the original texts were taken up a second time in the translations in the form of an ellipsis after a nominal quantifier (e.g. [9:3] and [9:4]).

(9:3) a) *Auf dem Spielplan stehen vor allem klassische Stücke, allerdings überwiegend in modernen Inszenierungen.*

(GO_TOU_006)

b) *The programme consists primarily of classical pieces, although most [] are given modern productions.* (ETRANS_TOU_006)

(9:4) a) *Das gotische Rathaus zählt zu Rostocks Kulturschätzen. Die Marienkirche ist auch dabei.* (GO_TOU_009)

b) *The Gothic townhall is one of Rostock's cultural treasures. Another [] is the Marienkirche.* (ETRANS_TOU_009)

We also find some additional nominal ellipses after possessive markers in the English translations (e.g. [9:5] and [9:6]) that do not have an exact structural equivalent in German.

(9:5) a) *Ich trat auf eine Platte [...] „Gehört deiner Mutter.“*

(GO_FICTION_006)

b) *I stepped on a record. [...] 'It's your mother's [].'*

(ETRANS_FICTION_006)

- (9:6) a) *Die Hüften, die sich darunter verbergen, können nicht nennenswert breiter als die Taille sein. Es sei denn, diese wäre insektenhaft schmal.* (GO_FICTION_004)
- b) *The hips concealed beneath it cannot be appreciably bigger than her waist – not unless the latter is as narrow as an insect's [].* (ETRANS_FICTION_004)

Most cohesive verbal/clausal ellipses correspond to similar structures in the translation (71 out of 90, i.e. 79%). The other cases were mainly translated by more explicit structures.

Figure 30 visualises these results and shows the percentages of cohesive ellipses that correspond to a similar type of ellipsis in the translation. We can see clearly in this figure that in translations from English into German, translators used structural equivalents for the majority of all cohesive ellipses whether they were nominal or verbal/clausal ellipses. From all cohesive ellipses that were used in the original texts, more than 60% were translated by the use of a cohesive ellipsis. In translations from German into English, most cohesive ellipses – if we look at all subtypes taken together – were also translated by similar structures, but we see striking differences between nominal and verbal/clausal ellipses here. Less than half of all cohesive nominal ellipses from German originals correspond to similar structures in the English translations, but the vast majority of

verbal/clausal ellipses – almost 80% – from German originals correspond to similar structures in the English translations.

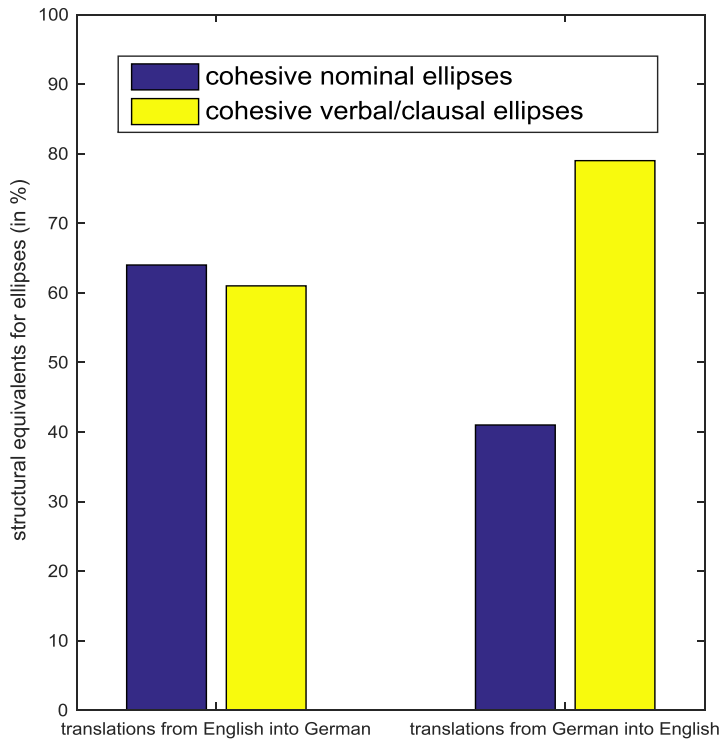


Figure 30: Percentage of cohesive ellipses from original texts that correspond to similar syntactic structures in the sentence-aligned translations

Another observation that emerged from the comparison of originals and translations is that antecedents as the first elements of cohesive chains often remain unchanged. If translators use category shifts between different types of cohesive devices, this usually affects anaphoric entities in our corpus data. Cohesive chains and category shifts of cohesive devices during translation processes are issues translators should be more aware of in order to make conscious decisions about their writing processes. (9:7) is

an example of a cohesive chain in a translated text with a different structure than in the original text.

(9:7) a) *There was not one way but many []. He followed first one [] and then the other [].* (EO_FICTION_002)

b) *Nicht ein Weg, sondern eine Vielzahl von Wegen lag vor ihm. Er folgte bald dem einen Pfad, bald dem anderen [].* (GTRANS_FICTION_002)

(9:7) contains three anaphoric nominal ellipses in the English original text that link to the antecedent noun ‘way’.²²⁸ In the German translation, the lexeme ‘Weg’ is the first word of the cohesive chain. After being reiterated, it is replaced by a more specific noun with a similar meaning (‘Pfad’) in the following sentence. Subsequently, an anaphoric nominal ellipsis is used, but this ellipsis is linked to the noun which was added in the translation and not to the same lexeme as in the original text. In this example, the translator avoided ellipsis-antecedent relations that stretch across sentence boundaries.

To sum up, we observed both explicitating and implicitating shifts in the translations, and in some cases entirely different sentence structures were used. Although a high proportion of ellipses from the original texts which were structurally modified during the translation process became

²²⁸ ‘one’/‘many’, ‘one’/‘other’ (and also the conjunctions ‘first’/‘then’) are semantically related in the English example and create lexical relations at the surface structure in the original text.

more explicit structures in the target texts, our data do not show a clear trend that would confirm the explicitation hypothesis in general. It is remarkable that, if we look at the total figures, most ellipses did not undergo any categorial shift during the translation process regardless of the translation direction involved. Like the almost identical frequencies of ellipses in the English and German original data discussed in Chapter 9.2, this observation equally reveals a close relatedness of the two languages in their systems of cohesion.

9.4 A comparison of written and spoken mode

From the spoken component, *ACADEMIC* and *INTERVIEW* were annotated for this analysis as these two spoken registers were included in the released version of *GECCO* at the time of the annotation. In order to have more spoken registers annotated than only two, internet forums texts (*FORUM*) from a later corpus version have been added to the annotated dataset. The corpus register *ACADEMIC* contains monologic, planned texts with many elements from the common language as specialised knowledge is popularised in these texts to a certain degree. The data in the register *INTERVIEW* come closest to everyday language and contain one-to-one, face-to-face conversations. The data in *FORUM* represent a more heterogeneous register. A variety of topics is covered depending on the specific text, and the texts contain narrative and descriptive passages as well as expressive dialogues. Although the *FORUM* texts often read as if they were spoken, they are a register between written and spoken language. Other spoken registers from a newer corpus version have not been added to the ellipsis annotation at this stage as they seem to be a slightly noisier dataset and potentially less comparable. There are many more contexts of spoken language use that are not represented in the corpus due to data scarcity and the difficulty to integrate a large amount of such data into a corpus. Due to the fact that there is no clear-cut boundary between spoken and written language in our corpus and that the size of the spoken data is smaller, this study has its main focus on general differences

between English and German and will only briefly discuss differences between written and spoken language.

The spoken corpus sections contain over 30 texts for both English and German respectively. Figure 9 in Chapter 8 has shown that the spoken corpus texts in GECCo are usually longer than the written texts, and Figure 17 (Chapter 9.2) suggested that there is a low to moderate degree of positive correlation between text length and number of cohesive ellipses.

Table 8 lists the frequencies of cohesive ellipses in the written and spoken corpus sections; Figure 31 visualises these results in a bar chart.

Subcorpus	tokens	all coh. ell. (abs.)	all coh. ell. (per 10,000 tokens)	coh. nom. ell. (abs.)	coh. nom. ell. (per 10,000 tokens)	coh. verbal / clausal ell. (abs.)	coh. verbal / clausal ell. (per 10,000 tokens)
EO_written	286,331	188	6.57	126	4.40	62	2.17
GO_written	288,490	187	6.48	97	3.36	90	3.12
EO_spoken	121,795	209	17.16	85	6.98	124	10.18
GO_spoken	125,537	210	16.73	100	7.97	110	8.76

Table 8: Absolute and normalised frequencies of all cohesive ellipses, cohesive nominal ellipses and cohesive verbal/clausal ellipses in written and spoken data in GECCo

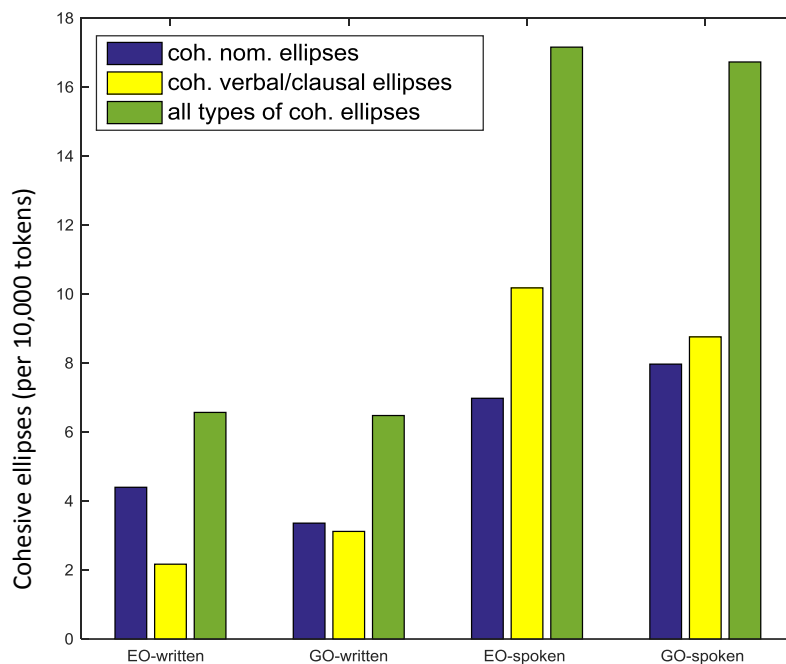


Figure 31: Normalised frequencies of cohesive nominal ellipses, cohesive verbal/clausal ellipses and sum of these types in written and spoken subcorpora

Regardless of a possible low to moderate influence of differences in text length on the results, we observed that the spoken corpus sections contain considerably more cohesive ellipses than the written ones. Figure 31 shows that all types of cohesive ellipses are more frequent in the spoken data, whether we look at them separately or at their sum (cf. also Figure 21 in Chapter 9.2 which showed the normalised frequencies of both types of ellipses in the texts in EO and GO in a stacked bar chart. A dotted line in that figure indicates the boundary between the written and spoken registers). Fictional texts (with many dialogues) in the written corpus sections also have high numbers of cohesive ellipses.

If we compare the total numbers of all types of cohesive ellipses, we see relatively similar results for the written subcorpora (ca. 6.5 cases per

10,000 tokens) and for the spoken subcorpora (ca. 17 cases per 10,000 tokens) respectively. Figure 31 also indicates that in the written English and German data, cohesive nominal ellipses are more frequent than cohesive verbal/clausal ellipses. In the spoken data, we can observe that all types of ellipses – nominal and verbal/clausal – are more frequent than in the written corpus section. Verbal/clausal ellipses are more frequent than nominal ellipses within the spoken data in both languages.

There is a slight difference between English and German with respect to the distribution of the ellipses subtypes: EO_written contains more nominal ellipses than GO_written; in the spoken data we can observe the opposite – the German data contain more nominal ellipses. EO_written, on the other hand, is characterised by fewer cohesive verbal/clausal ellipses than GO_written. Again, the opposite effect can be observed in the spoken data – in EO_spoken, there are more cohesive verbal/clausal ellipses than in the German spoken data.²²⁹

Figure 32 visualises the spread and differences of the sample distributions for the spoken subcorpora via comparative boxplots.

²²⁹ As the spoken data contain fewer texts, we did not run significance tests for the comparison of written and spoken data. Nevertheless, we should keep in mind that the statements on the written data are probably more representative and less likely to be explained by chance.

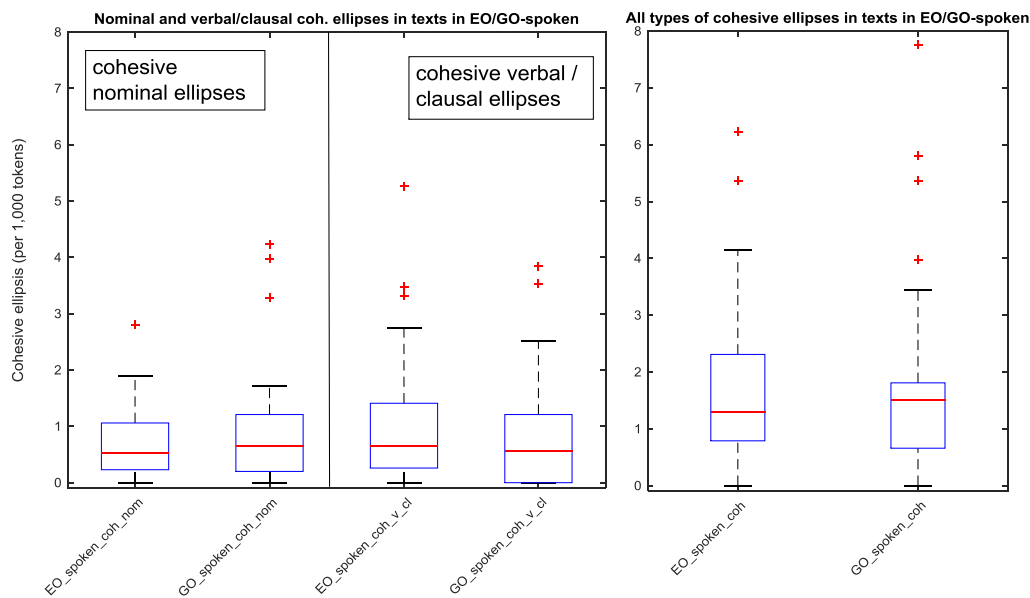


Figure 32: Normalised frequencies of cohesive ellipses in spoken corpus sections

The first subplot shows the distribution of nominal and verbal/clausal ellipses in EO_spoken and GO_spoken separately. The second subplot is a visualization of the distribution of all cohesive ellipses in the spoken corpus texts (cf. also the boxplots in Figure 28 on cohesive ellipses in EO_written and GO_written that have much smaller interquartile ranges and low medium values, but nevertheless some similarly high values for outliers). The percentiles and the interquartile range are rather similar for the English and German spoken data in the first subplot. In the second subplot, the median is higher and we find more outliers in the German

spoken data, but the interquartile range is larger in the English spoken data.

The main observation in the comparison of written and spoken data is that ellipses as cohesive devices are generally more frequent in spoken texts in both English and German. The two languages show similarities and comparable frequencies for the distributions of cohesive ellipses and their subtypes. So far the analyses may have suggested that ellipses are only used occasionally to contribute to the cohesiveness of texts, but the spoken data present us with a different picture. Particularly, verbal/clausal ellipses seem to be not unusual as cohesive devices in spoken language use. Cohesive nominal ellipses are also more frequent in spoken language, but this type of ellipses is not the most frequent ellipsis type in the spoken data. Nevertheless, it is the most dominant ellipsis type in the written corpus sections.

9.5 Register variation

This section gives an overview on variation between the different corpus registers and focuses for this analysis on the corpus sections of non-translated texts. Figure 21 in Chapter 9.2 has already shown the frequencies of both types of ellipses in the texts in GECCO-EO and GO with lines indicating register boundaries. That figure has already suggested that texts belonging to the same corpus register behave rather similar even across languages. Table 9 lists the absolute and normalised frequencies of nominal and verbal/clausal cohesive ellipses and their sum for each corpus register.

Register	tokens	nominal coh. ell. (abs.)	nominal coh. ell. (per 10,000 tokens)	verbal / clausal cohesive ell. (abs.)	verbal / clausal cohesive (per 10,000 tokens)	all cohesive ell. (abs.)	all cohesive ell. (per 10,000 tokens)
English							
EO_ESSAY	34,998	14	4.00	4	1.14	18	5.14
EO_FICTION	36,996	44	11.89	30	8.11	74	20.00
EO_INSTR	36,167	3	0.83	2	0.55	5	1.38
EO_POPSCI	35,148	10	2.85	4	1.14	14	3.98
EO_SHARE	35,824	11	3.07	5	1.40	16	4.47
EO_SPEECH	35,062	12	3.42	6	1.71	18	5.13
EO_TOU	35,907	26	7.24	4	1.11	30	8.35
EO_WEB	36,119	6	1.66	7	1.94	13	3.60
EO_ACADEMIC	40,559	29	7.15	28	6.90	57	14.05
EO_FORUM	43,338	26	6.00	29	6.69	55	12.69
EO_INTERVIEW	37,898	30	7.92	67	17.68	97	25.60
German							
GO_ESSAY	35,668	18	5.05	3	0.84	21	5.89
GO_FICTION	36,778	26	7.07	70	19.03	96	26.1
GO_INSTR	36,880	7	1.90	4	1.08	11	2.98
GO_POPSCI	36,177	10	2.76	8	2.21	18	4.98
GO_SHARE	35,235	2	0.57	0	0	2	0.57
GO_SPEECH	35,399	11	3.11	3	0.85	14	3.95
GO_TOU	36,574	17	4.65	1	0.27	18	4.92
GO_WEB	35,779	6	1.68	1	0.28	7	1.96
GO_ACADEMIC	43,703	40	9.15	28	6.41	68	15.56
GO_FORUM	41,636	39	9.37	43	10.33	82	19.69
GO_INTERVIEW	40,198	31	7.71	29	7.21	60	14.93

Table 9: Absolute and normalised frequencies of nominal and verbal/clausal cohesive ellipses and their sum for each corpus register

The following pie charts visualise this information and illustrate how the different ellipsis types are distributed across the individual 11 registers in the English and German subcorpora of original texts. Each sector or segment represents a certain proportion or percentage of the total. The pie charts in Figure 33 show the percentages of all cohesive ellipses per

register in the English and German data respectively (based on normalised figures) and illustrate which registers are characterised by high or low proportions of the total number of ellipses found in the corpus.

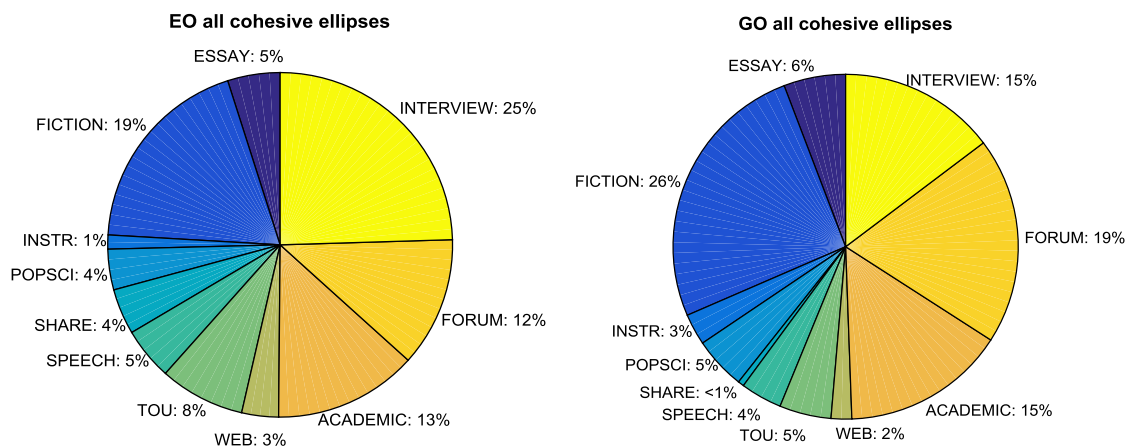


Figure 33: Proportions of cohesive ellipses in English and German data

The registers that contain high proportions of cohesive ellipses in both languages are ACADEMIC, INTERVIEW and FORUM from the spoken corpus sections and FICTION from the written section. Taken together, these four sections contain about 70% of all cohesive ellipses in EO and 75% of all cohesive ellipses in GO. Among the written data, ellipses are particularly rare in the three registers of INSTR, SHARE and WEB. Each of these registers accounts for less than 5% of all cohesive ellipses in both languages. Taken together, they only contain about 6 to 8% of all cohesive ellipses.

Figures 34 and 35 visualise these figures separately for the individual subtypes of cohesive ellipses.

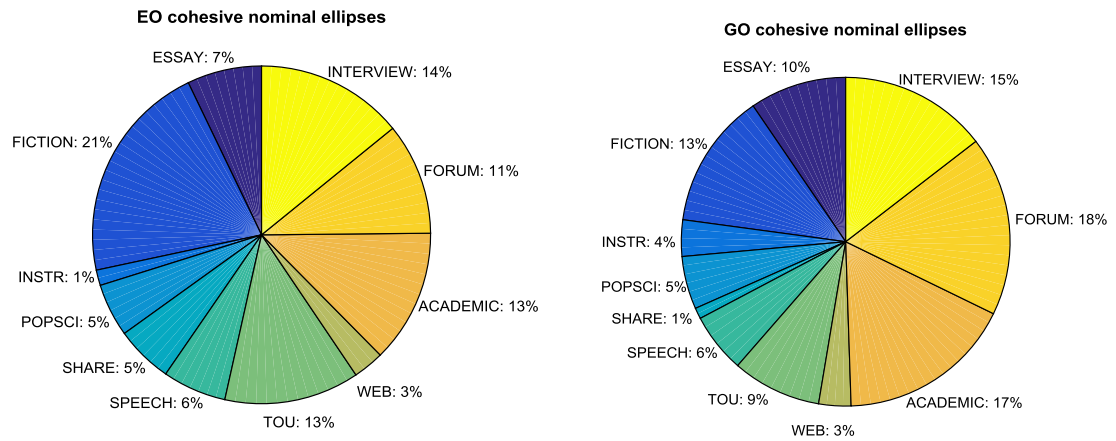


Figure 34: Proportions of nominal cohesive ellipses in English and German data

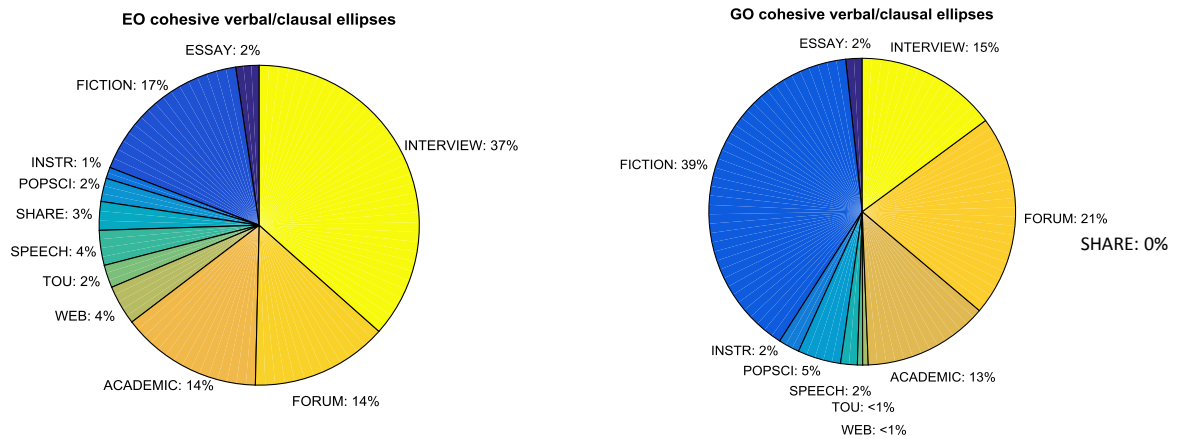


Figure 35: Proportions of cohesive verbal/clausal ellipses in English and German data

For nominal ellipses, the distribution is rather similar in English and German. Many ellipses of this type can be found in the spoken sections and in FICTION, although German has more nominal ellipses in the spoken corpus sections and fewer in FICTION compared to English. Tourism texts and political essays also have quite high proportions of cohesive nominal ellipsis, as these registers are rich in noun phrases with modifiers combined with syntactic parallelisms and semantic contrasts which trigger elliptical structures.

With cohesive verbal/clausal ellipses we find even more striking differences between the corpus registers. The registers from the spoken sections and FICTION include almost 82% of these ellipses in EO and 88% in GO. The written registers apart from FICTION have relatively similar proportions in English (mostly between 2 and 4%). In German, these registers have maximally 2%; only POPSCI has 5%. In both languages, there is one in 11 registers that contains almost 40% of all cases of cohesive verbal/clausal ellipses – INTERVIEW in English and FICTION in German. As the corpus includes only about 10 texts for these registers, this analysis gives only a rough idea of the distribution of ellipses in different registers. I do not claim that these results are statistically significant and the results in Annex 1 also indicate that individual texts in INTERVIEW and FICTION can have a large effect on the results: In English interviews, some texts have 0 cases and others more than 15. In German fictional texts, 50% of the texts have zero or one occurrence but some have between 10 and 20 or even 40 cases.

For comparative purposes, Table 10 shows the absolute and normalised frequencies of non-cohesive ellipses and other fragments.

Register	non-coh. nominal ell. (abs.)	non-coh. nominal (per 10,000 tokens)	non-coh. verbal/clausal (abs.)	non-coh. verbal/clausal (per 10,000 tokens)	all non-coh. (abs.)	all non-coh. (per 10,000 tokens)	other fragments (abs.)	other fragments (per 10,000 tokens)
English								
EO_ESSAY	15	4.29	6	171	21	6	133	38
EO_FICTION	23	6.22	32	8.65	55	14.87	208	56.22
EO_INSTR	1	0.28	4	1.11	5	1.38	356	98.43
EO_POPSCI	15	4.27	11	3.13	26	7.40	78	22.19
EO_SHARE	11	3.07	6	1.67	17	4.75	135	37.68
EO_SPEECH	10	2.85	14	3.99	24	6.85	72	20.54
EO_TOU	15	4.18	5	1.39	20	5.57	336	93.58
EO_WEB	6	1.66	12	3.32	18	4.98	350	96.90
EO_ACADEMIC	10	2.47	9	2.22	19	4.68	11	2.71
EO_FORUM	17	3.92	25	5.77	42	9.69	380	87.68
EO_INTERVIEW	2	0.53	23	6.07	25	6.60	109	28.76
German								
GO_ESSAY	9	2.52	15	4.21	24	6.73	227	63.64
GO_FICTION	23	6.25	16	4.35	39	10.60	361	98.16
GO_INSTR	6	1.63	5	1.36	11	2.98	962	260.85
GO_POPSCI	19	5.25	7	1.93	26	7.19	91	25.15
GO_SHARE	3	0.85	1	0.28	4	1.14	146	41.44
GO_SPEECH	12	3.39	7	1.98	19	5.37	161	45.48
GO_TOU	4	1.09	21	5.74	25	6.84	404	110.46
GO_WEB	9	2.52	6	1.68	15	4.19	342	95.59
GO_ACADEMIC	17	3.89	11	2.52	28	6.41	94	21.51
GO_FORUM	22	5.28	22	5.28	44	10.57	283	67.97
GO_INTERVIEW	2	0.50	5	1.24	7	1.74	110	27.36

Table 10: Absolute and normalised frequencies of non-cohesive ellipses and fragments for each corpus register

Due to space constraints, not all graphical representations of the results that were generated for visualising register differences can be presented here, but it is worth pointing out that *non-cohesive* ellipses are distributed

much more evenly across the different registers in both English and German. There are fewer non-cohesive ellipses than cohesive ellipses if we compare the absolute frequencies. Nevertheless, the proportions in percentages of non-cohesive ellipses do not differ strongly between written and spoken registers. Moreover, in both the written and spoken data, we find registers with high or low proportions of non-cohesive ellipses. POSPCI for instance is a written registers that contains relatively high proportions of non-cohesive ellipses while INTERVIEW as a spoken register contains low proportions of non-cohesive ellipses. INSTR is a written registers that contains low proportions of non-cohesive ellipses whereas FORUM from the spoken corpus section contains high proportions of non-cohesive ellipses. I had assumed that all types of ellipses are indicators of spoken, dialogic and addressee-oriented registers and occur more frequently in such registers in both English and German, but with non-cohesive ellipses we do not see a certain trend that would confirm this hypothesis.

Figure 36 visualises the distributions of other fragments across the English and German registers.

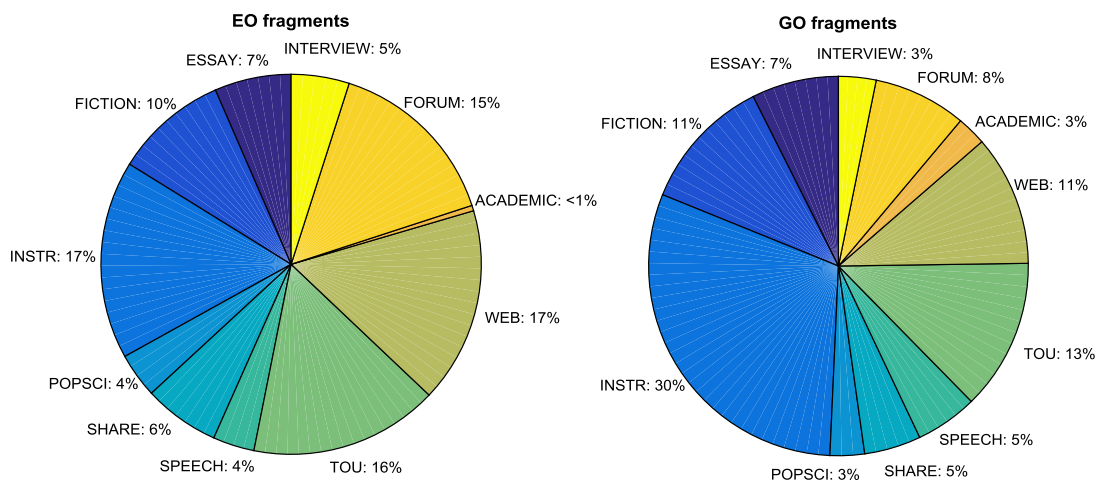


Figure 36: Proportions of fragments in English and German data

Here again, we do not see a clear distinction between the written and spoken registers. Although the results for ACADEMIC and INTERVIEW are particularly low, we find relatively high frequencies in FORUM. Among the registers from the written corpus sections, there are some with either low or higher frequencies in both languages. In Chapter 9.2, it has been demonstrated that fragments are considerably less frequent in English than in German and that there is a higher variability in the German data. Nevertheless, if we compare English and German, the distribution across the registers that we see in Figure 36 is rather similar. The high results for fragments in INSTR and WEB in both languages are noticeable as these registers had very low frequencies for all types of ellipses. In Chapter 9.2, it has been shown that I did not find any general relationship between the number of ellipses that are used as cohesive devices in the texts in EO and GO and the number of other fragments in these texts (Figure 26).

Fragments occur independently of ellipsis frequencies in texts that typically contain many non-clausal units, lists, headlines and keywords such as instruction manuals, tourism leaflets or certain text types from online sources. Due to the heterogeneous mix of different text styles in FICTION and FORUM, these registers contain relatively high frequencies of all types of structures that I annotated. Both contain elements from written and spoken language, dialogues as well as narrative passages and monologues.

Figure 37 illustrates the distributions of several types of cohesive devices per language and register²³⁰ in the GECCo corpus. In this figure, the percentages of personal, demonstrative and comparative reference, substitution, ellipses (nominal, verbal, clausal taken together), conjunction and general nouns as the one subtype of lexical cohesion are visualised. Lexical cohesion is the most frequent type among all cohesive devices, but not all corpus data and subcategories have been annotated for this category so far.

²³⁰ without FORUM

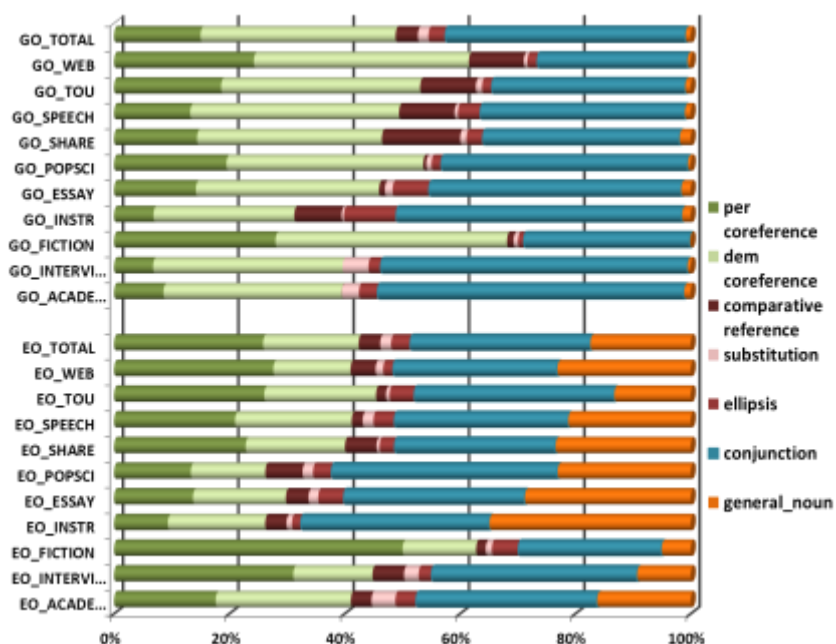


Figure 37: Distributions of different cohesive devices per language and register (from Kunz et al., forthcoming, a)

With reference to the topic of grammatical ellipsis-antecedent mismatches raised in Chapter 3.3, I systematically searched through the data and found that both English and German contain relatively few mismatches in ellipsis-antecedent relations. Several mismatches in German involve grammatical case mismatches in nominal ellipses, which is nothing unusual. Otherwise English and German mismatches in our data typically involve nominal singular or plural antecedents and ellipsis remnants that could be filled with another form of the noun (9:8-10). Grammatical mismatches typically occur in contexts of contrasts between the elements in the ellipsis remnant and the antecedent phrase, which explains that an ellipsis can be filled with the same lexeme, although this

lexeme does not always match with its antecedent in every grammatical aspect.

(9:8) *This is a combination of two recipes; one [] was a pound cake and the other [] was a yellow cake. (EO_FORUM_010)*

(9:9) *1999 was an outstanding year, but the past five [] have been great as well, as have the past 20 []. (EO_SHARE_004)*

(9:10) *Die Agenda ist nur der erste Schritt. Es müssen andere [] folgen. (GO_ESSAY_005)*

As predicted in Chapter 5, such cases most often occur in nominal ellipses after numerals or quantifiers where one element is contrasted with another or with several elements (cf. also examples 5:37-38 above). They also occur in contexts involving comparative forms of adjectives. Table 11 lists the absolute numbers of contexts per corpus register in which singular-plural mismatches occurred between a nominal ellipsis and its antecedent.

	Number mismatches between nominal ellipsis and antecedent (EO)	Number mismatches between nominal ellipsis and antecedent (GO)	Number mismatches between nominal ellipsis and antecedent (EO)	Number mismatches between nominal ellipsis and antecedent (GO)
	cohesive ellipses		non-cohesive ellipses	
ESSAY	3	2	1	1
FICTION	3	4	2	1
INSTR	0	1	0	0
POPSCI	0	1	3	3
SHARE	2	0	2	0
SPEECH	1	1	2	1
TOU	7	3	4	0
WEB	0	1	1	1
ACADEMIC	7	10	1	2
FORUM	9	6	2	1
INTERVIEW	2	4	2	1

Table 11: Number mismatches between nominal ellipses and antecedents

We find most of these number mismatches for cohesive nominal ellipses in the spoken corpus sections and in English tourism leaflets. Otherwise this type of mismatches seems to be rare in written texts for cohesive ellipses and in all registers for non-cohesive ellipses. Initially, I assumed that ellipsis-antecedent mismatches can be expected to occur within clause complexes and do not span longer distances. Nevertheless, distance is probably not the most important criterion. Topics involving contrast in spoken and addressee-oriented registers seem to be a more important factor.

Grammatical mismatches with regard to person, number, voice or tense of in verbal/clausal ellipses occur even more rarely in our data. They can sometimes be found in informal registers such as FORUM (9:11) or in creative texts such as FICTION (cf. example [3:39] above).

(9:11) *I'm supposed to avoid both and think I am [].*

(EO_FORUM_002)

Verbal mismatches are slightly more frequent in English compared to German and tend to occur when the ellipsis remnant involves an auxiliary or a modal verb or in question-answer pairs in dialogues where, for instance, *I* becomes *you* and vice versa.

In summary of this chapter, it is evident from the quantitative corpus analysis that cohesive ellipsis-antecedent relations are not extremely frequent, but they occur occasionally and contribute to the textuality of written and oral texts. In most cases, cohesive elliptical structures involve a verbatim recoverability of the missing elements. Statistical analyses of the data have been provided that seem to confirm some of the initial hypotheses and shed light on areas where no clear hypotheses and no previous quantitative studies exist. This study has investigated how frequent cohesive ellipses and its subtypes are in the dataset under analysis and where differences and similarities lie between languages, registers, spoken and written modes and between originals and translations.

The study confirmed that cohesive ellipses occur a handful of times in most texts in both languages. In some texts, ellipses appear abundantly while in others, they do not occur at all. I expected more ellipses of the cohesive and the non-cohesive type in English compared to German, but the overall frequencies were almost identical in both languages. German only has more fragments as irregular syntactic structures than English. The

translation data contained many cases where abridged structures became more explicit by adding lexical material in the translation process, but I did not find any evidence that would clearly support the explicitation hypothesis. In general, elliptical patterns in the translations reflect features of the source language texts. I found cross-linguistic similarities between the registers regardless of the language involved as well as striking differences between written and spoken data. Cohesive ellipses indeed occur more frequently spoken registers in both English and German. Finally, the data indicate that German is characterised by a stronger registerial differentiation and stronger distinctions along the dimensions of written vs. spoken and formal vs. colloquial style than English. The smaller the respective dataset the more cautious our conclusions on the basis of our samples in the subcorpora have to be with respect to the underlying ‘population’ as a whole.

The corpus examples of ellipses demonstrated that, although ellipses can cross sentential and clausal boundaries, they are not completely independent of the grammatical structure of the text and usually do not span long distances. They often occur in pairs of sentences with parallelisms involving complementary relationships, contrasting and similarity relationships or sequential relationships. In ellipsis contexts, similarly structured elements and some morphosyntactic information are reiterated where one element can be left out. It can be concluded that ellipsis sites indeed induce structural priming effects as suggested by Xiang et al. (2014).

The structures that were covered in this study may seem diverse to a certain extent as some are omissions of a single word within a complex sentence while others involve omissions of complex structures so that only one or a few words are left as remnants. Nevertheless, it has been demonstrated that all these omissions in both English and German can be subsumed under the categories of nominal, verbal and clausal ellipses.

10. Conclusions and outlook

One of the primary goals of this dissertation was the conceptual clarification of ellipses as cohesive devices and the operationalisation of this concept for a contrastive study. Assuming that there is unpronounced syntactic structure in ellipsis sites, I examined ellipsis-antecedent relations that contribute to the cohesiveness of texts. The broad view on cohesive ellipses as implicit or unsaid textual elements in the previous literature on discourse relations has been narrowed down to a more precise definition, and it has thereby been confined to reasonable limits. Our work provides a detailed annotation scheme of ellipses as cohesive devices which covers different types of nominal, verbal and clausal ellipses. These range from the omission of single words within various types of phrasal and clausal structures that result in ellipsis remnants embedded in sentences with different levels of complexity to the omission of larger groups of words and constituents that result in short remnant structures. This study has emphasised the value of corpus studies as the patterns and distribution of elliptical structures that were found in the data partly deviate from assumptions and standard examples from the theoretical literature.

This dissertation contributes to the field of English-German contrastive linguistics with a focus on texts as communication objects. The corpus analysis that has been conducted illustrates the different ellipsis types with authentic examples embedded in different registerial contexts in both English and German. The study has shown how particular types of ellipses

as textual omissions establish textual relations that contribute to the cohesiveness and coherence of texts. The quantitative results of the corpus study have provided answers to our research questions and have shown how cohesive ellipses are distributed in English and German texts. They also show how variables such as register, the distinction between written and spoken language and production mode (original texts vs. translations) affect the use of these structures. Cohesive ellipses are relatively rare in our corpus data in both languages. Nevertheless, they are used with similar frequency distributions in comparable sections of the corpus. Apparently, textual cohesion is an area where English and German employ rather similar strategies with similar frequencies.

The annotations that were created and the extracted patterns can serve as a basis for similar annotations in a larger mono- or bilingual corpus in the future to obtain higher absolute frequencies of elliptical structures. The most frequent and most relevant grammatical patterns of ellipsis contexts that were identified in each register can help to improve semi-automatic annotation methods and to restrict automatic queries to the most typical patterns for a given text type or corpus. It is left for future empirical work to determine whether the methods of this work can also be applied to other languages, particularly to other Germanic languages or to Romance languages that can be expected to use similar strategies to establish textual cohesion.

Another aspect that could stimulate further research is the study of intonation, stress and rhythm of speech. It would be interesting to examine

the prosodic structure of different types of ellipsis-antecedent relations and to investigate the interplay of our ellipsis categories and focus structure. The contexts containing ellipses in our data could serve as a database for prosody research. For the spoken corpus data, the original audio files are also available.

Cases of subject ellipsis in coordinated structures were not the main focus of my analysis of ellipses contributing to textual cohesion. Although they are a frequent sentence-internal linguistic phenomenon, I decided to exclude them entirely from my study with the intention to address this topic in future research.

The results of this study are relevant to the questions theoretical and applied linguistics seeks to answer, but they have several practical implications for foreign language pedagogy and translator training as well. In general, English and German seem to behave rather similarly when it comes to the use of ellipses in their function as cohesive devices. Nevertheless, it is necessary to be aware of the functions of ellipses as omissions within and across sentence boundaries and the functions of other types of fragments or independent non-clausal units in order to be able to use these structures appropriately and in language- and register-typical frequencies. The decision to use an ellipsis, substitution, co-reference items or lexical cohesion to create textual cohesion influences the structure of lexical chains in texts.

The different types of cohesive devices that have been annotated can now be queried in the GECCo corpus. The textual annotations of ellipses

are freely accessible to researchers and students via corpus queries. Additionally, all corpus examples of endophoric ellipses have been collected and listed. These lists also include information on the respective text type, the exact subtype of ellipsis and the strategy translators used to achieve an equivalent effect in the translation or to achieve a different effect if they preferred or were obliged to use different structures. Exercises and worksheets on the use of ellipses, lexical cohesion and substitution have been developed to be used in translator training and linguistic seminars. These teaching materials contain numerous examples of ellipses in original and translated texts with comments on the type and function of the ellipsis and discussions of translation strategies. In our materials, ellipsis as a grammatical category is illustrated by examples from corpus texts, their commented translations and paraphrased variants as alternative translations. The pedagogical implications of the GECCo project are discussed in more detail in Menzel (2016c). The analysis of these data can be a useful method in foreign language pedagogy and translator training to help learners develop textual competence in both their mother tongue and the foreign language.

11. Zusammenfassung der Dissertation (Summary in German)

Gegenstand dieser Dissertation ist eine kontrastive Analyse von elliptischen Strukturen im Englischen und Deutschen, die als potentielle Textverknüpfungsmittel mit bereits vorerwähnten Textelementen in Verbindung stehen können. Bei Ellipsen handelt es sich um ein spezielles Kohäsionsmittel, das sich von den anderen lexikalischen und grammatischen Mitteln zur Textverknüpfung dadurch unterscheidet, dass eine Relation nicht zwischen zwei eigentlichen Textelementen an der Textoberfläche hergestellt wird, sondern zwischen einem Antezedens im Text und einer darauf folgenden Auslassung eines grammatisch-syntaktisch notwendigen Satz- oder Phrasenbestandteils. Elemente, die im Text ausgespart und somit in der Oberflächenstruktur getilgt wurden, können unterschiedlichen Wort- und Phrasenarten angehören, und sind – im Unterschied zu allen anderen Kohäsionsmitteln – nur in der syntaktischen Tiefenstruktur vorhanden. Ohne Bezug auf den Kontext, in den eine Auslassung eingebettet ist, ist eine Analyse ihrer Struktur nicht möglich. Ellipsen gelten bisher als besonders schwierig zu beschreibende Strukturen, die mit empirischen Methoden schwer erfassbar sind.

Ein wesentliches Ergebnis dieser Dissertation besteht darin, dass sie den Begriff der kohäsiven Ellipse in Abgrenzung von anderen Phänomenen präzisiert und operationalisierbar macht und ihn dann für eine konkrete sprachkontrastive Analyse nutzt. Nicht alle endophorischen Ellipsen sollten als Kohäsionsmittel aufgefasst werden, sondern nur solche, die

satz- oder teilsatzübergreifend verwendet werden. Ausgeschlossen werden Fälle, die ausschließlich in Koordinationen möglich sind und dabei nicht zur Verknüpfung von über die koordinierte Struktur hinausgehenden Textpassagen geeignet sind. Zu den nicht-kohäsiven Ellipsen zählen auch exophorische, situationsdeiktische Beispiele und kontextunabhängige, konventionalisierte Ellipsen. Von elliptischen Strukturen insgesamt sind andere Arten von Satzfragmenten abzugrenzen.

Diese Arbeit leistet einen Beitrag für die kontrastive Beschreibung der englischen und deutschen Grammatik mit einem Schwerpunkt auf Texten als Kommunikationsinstrumenten. Es wird hierbei auf eine korpusbasierte Analyse sowie auf nicht-konstruierte, authentische Beispiele der untersuchten Textverknüpfungsmittel, die in einen sprachlichen Kontext eingebettet sind, Wert gelegt. Hierzu wurden verschiedene Subtypen elliptischer Strukturen und ihre gegebenenfalls vorhandenen textuellen Bezugselemente in einem bilingualen Korpus annotiert und hinsichtlich ihrer Häufigkeiten ausgewertet. Es wird der Frage nachgegangen, wie sich Englisch und Deutsch im Hinblick auf den Gebrauch von kohäsiven Ellipsen unterscheiden oder ähneln, aber auch welche Rolle verschiedene Textproduktionstypen (Originale vs. Übersetzungen) und Unterschiede zwischen verschiedenen Textsorten und gesprochensprachlich oder schriftsprachlich konzipierten Texten spielen. Nach der Entwicklung eines detaillierten Annotationsschemas und der Annotation des GECCo-Korpus wird untersucht, wie häufig kohäsive Ellipsen und verschiedene Subtypen dieser Kategorie in den verwendeten Korpusdaten vorkommen.

Im Folgenden wird ein Überblick über die Vorgehensweise und den Aufbau der Arbeit gegeben. Das Einleitungskapitel stellt die Thematik und die Fragestellungen der Arbeit vor. Des Weiteren werden die Hypothesen der im Rahmen der Arbeit durchgeführten empirischen Untersuchung aufgeführt. Es wird erörtert, welche Häufigkeitsverteilungen für Ellipsen in den englischen und deutschen Korpusdaten erwartet werden in Abhängigkeit von der Sprache, den jeweiligen Textproduktionstypen (Originaltexte vs. Übersetzungen), den einzelnen Korpusregistern und den generellen Unterschieden zwischen geschriebenen und gesprochensprachlichen Textsorten. Zu den konkreten Fragestellungen der Arbeit, die bisher in dieser Form nicht empirisch untersucht wurden, lassen sich aus der sprachwissenschaftlichen Literatur keine Aussagen finden und nur indirekte Hypothesen ableiten. Es kann vermutet werden, dass im Englischen insgesamt mehr Auslassungen und daher auch mehr kohäsive Ellipsen als im Deutschen verwendet werden, wobei hier aber zwischen den verschiedenen Ellipsentypen differenziert werden muss. In beiden Sprachen sind alle untersuchten Ellipsenarten prinzipiell möglich und werden mit gleichen Funktionen verwendet. Bestimmte Einzelstrukturen, die unter die festgelegten Ellipsenkategorien fallen, sind jedoch aus grammatischen Gründen in jeweils einer der beiden Sprachen nur eingeschränkt möglich. Bei den Übersetzungen wird vermutet, dass in diesen ähnliche Ellipsenhäufigkeiten zu finden sind, wie in den zugehörigen Originaltexten der anderen Sprache. Möglicherweise lässt sich auch ein Explizierungseffekt feststellen, sodass anstelle von Ellipsen

auch bevorzugt explizitere Strukturen zum Erzeugen von Textkohäsion verwendet werden. Außerdem kann erwartet werden, dass die Häufigkeiten von Ellipsen in beiden Sprachen stark von der Textsorte abhängen und dass es große Unterschiede zwischen geschriebener und gesprochener Sprache sowohl im Englischen als auch im Deutschen gibt.

In Kapitel 2 werden einige wichtige Aspekte der Grammatikgeschichte des Begriffs diskutiert, die sich bis in die Antike zurückverfolgen lässt. Frühe Grammatikographen haben Begrifflichkeiten geprägt, die teilweise aus philosophischen und rhetorischen Diskursen hervorgegangen sind, und Diskussionen angestoßen, die unsere heutigen Grammatikmodelle und unser Grammatikverständnis nicht unerheblich beeinflusst haben.

Kapitel 3 behandelt mehrere Diskussionspunkte, die in der linguistischen Forschung im Zusammenhang mit Ellipsen in jüngster Zeit kontrovers diskutiert wurden. Dazu gehören die Frage der Textgliederung und Einheitenbildung in der geschriebenen im Gegensatz zur gesprochenen Sprache, die Schnittstelle zwischen Syntax, Semantik und Pragmatik, die Frage nach der Identität zwischen den elidierten Elementen und ihren Antezedenzen sowie Skopus- und Ambiguitätsprobleme bei Ellipsen.

Kapitel 4 beschäftigt sich mit verschiedenen Möglichkeiten, Subkategorien für Ellipsen zu bilden und mit Typologien und Taxonomien, die in der Literatur bisher zur Unterscheidung von unterschiedlichen Ellipsenarten vorgeschlagen wurden. In diesem Kapitel wird aufgezeigt, warum viele existierende Beschreibungsansätze für

Ellipsen problematisch für empirische Arbeiten sind, da in ihnen Kategorien nicht scharf voneinander abgegrenzt werden oder als graduelle Phänomene behandelt werden, deren Grenzen fließend sind. Die gewählten Kategorisierungsdimensionen und Selektionskriterien sowie die Entscheidung, wie eng oder breit sind Kategorien gefasst werden, haben einen entscheidenden Einfluss auf quantitative Untersuchungen anhand von Korpusdaten. Die Vielzahl elliptischer und fragmentarischer Strukturen, die in der Vergangenheit unter den Ellipsenbegriff subsumiert wurde, sind zudem bisher kaum im Zusammenhang mit ihrer möglichen Funktion als Kohäsionsmittel untersucht worden. Es standen meist isolierte oder satzinterne Phänomene bei der Beschreibung von Ellipsentypen im Mittelpunkt. Für die vorliegende Arbeit ist es notwendig, die zahlreichen Ellipsensubtypen zu allgemeineren, abstrakteren Kategorien zusammenzufassen, um die unterschiedlichen Auslassungsmöglichkeiten in den verschiedenen Registern der verwendeten englischen und deutschen Korpusdaten vergleichen zu können. Kapitel 4 bietet auch einen Überblick über die Einordnung des Ellipsenthemas in den systemisch funktionalen Ansatz und zeigt Bereiche auf, die in der bisherigen Literatur noch nicht klar genug herausgearbeitet wurden.

Kapitel 5 widmet sich ausführlich dem in dieser Arbeit entwickelten Annotationsschema für kohäsive Ellipsen. Es beschreibt detailliert, welche Fälle im Englischen und Deutschen unter nominale, verbale und klausale Ellipsen fallen und wodurch sich kohäsiv verwendete Ellipsen von nicht-

kohäsiv verwendeten Ellipsen und anderen, den Ellipsen sehr ähnlichen Phänomenen, wie z.B. Substitution oder Right-Node-Raising unterscheiden. In diesem Kapitel werden die einzelnen Kategorien des Annotationsschemas anhand zahlreicher authentischer Korpusbeispielen verdeutlicht, um eine Vielzahl möglicher Kontexte zu beschreiben, in denen Ellipsen im tatsächlichen Sprachgebrauch in verschiedenen Textsorten vorkommen können. Ziel des Annotationsschemas ist es, jeden der vielfältigen Fälle, die im Korpus vorkommen, möglichst eindeutig einer Kategorie zuzuordnen. In diesem Modell werden Aspekte aus unterschiedlichen existierenden Ellipsenbeschreibungen berücksichtigt und in eine konsistente Beschreibung aller Ellipsenarten integriert. Theoretische Überlegungen werden hierbei mit den untersuchten Korpusdaten verknüpft.

In Kapitel 6 wird darauf hingewiesen, dass einige Arten von Fragmenten und nicht-satzförmigen Äußerungen nicht notwendigerweise durch Reduktion oder Auslassung zu erklären sind und daher nicht mit den in Kapitel 5 beschriebenen Ellipsen verwechselt werden sollten. Nicht alle Strategien der Sprachökonomie tragen zur verstärkten Textkohäsion bei. In diesem Kapitel werden einige Phänomene kurz erläutert, die manchmal in der Literatur ebenfalls unter den Ellipsenbegriff fallen, die hier aber als konzeptuell unterschiedliche Strukturen aufgefasst werden. In den Korpusdaten wurden sie annotiert, um sie von den Ellipsen mit Textverknüpfungspotential oder -charakter abzugrenzen.

Kapitel 7 vergleicht die Funktion von kohäsiven Ellipsen und anderen

Kohäsionsmitteln. Im Vergleich zu anderen Textverknüpfungsmitteln werden kohäsive Ellipsen sparsamer, dafür aber meist gleichzeitig als Stilmittel mit bestimmten Funktionen verwendet. Zu den Funktionen von Ellipsen gehört auch, Wortwiederholungen zu vermeiden und bestimmte Konstituenten besser hervorzuheben. Für kohäsive Ellipsen existieren verschiedene Alternativen unter den Kohäsionsmitteln. Eine Ellipse kann insbesondere durch die Verwendung von lexikalischen Kohäsionsmitteln oder Pro-Formen, die als Substitute eingesetzt werden, vermieden werden. Lexikalischen Kohäsionsmittel dienen beispielsweise dazu, einen vielfältigen Wortschatz in einem Text einzusetzen oder Wiederholungen gezielt einzusetzen. Bei Substituten handelt es sich um eine kleine Gruppe grammatischer Elemente mit wenig semantischem Inhalt.

Kapitel 8 fasst bisherige korpusbasierte Studien zu Ellipsen zusammen. Bisher durchgeführte Untersuchungen zu verschiedenen, als Ellipsen bezeichneten Strukturen, basieren auf unterschiedlichen Ansätzen und Interessenschwerpunkten und widmen sich meist einem äußerst eingegrenzten Phänomenbereich oder relativ kleinen, monolingualen Korpora. Sofern elliptischen Strukturen in großen existierenden Korpora bisher untersucht wurden, war der Schwerpunkt nicht auf quantitativen Fragestellungen, da bei automatischen Abfragen Trefferquote und Genauigkeit nicht sehr hoch waren und die Korpusdaten lediglich dazu dienten, eine Auswahl authentischer Belege zu finden. In diesem Kapitel werden auch das in dieser Studie verwendete Korpus sowie die zur Kompilierung, Annotation, Abfrage und Extraktion verwendeten Tools

und Methoden beschrieben. Das verwendete GECCo-Korpus besteht aus englischen und deutschen Originaltexten aus verschiedenen geschriebenen und gesprochenen Registern (mehr als 1,5 Mio. Tokens), wobei für die geschriebenen, publizierten Texte (politische Aufsätze, Belletristik, Bedienungsanleitungen, populärwissenschaftliche Texte, Aktionärsbriefe, vorformulierte politische Reden, Tourismusprospekte und Webseiten von Firmen und Organisationen) auch ihre Übersetzungen in die jeweilige andere Sprache vorliegen. In dem Korpus sind inzwischen neben kohäsiv verwendete Ellipsen auch andere Arten von Kohäsionsmitteln annotiert (Referenz, Substitution, Konnektoren und lexikalische Kohäsion). Für die vorliegende Untersuchung zu Ellipsen wurden alle englischen und deutschen geschriebensprachlichen Texte untersucht. Als gesprochensprachliche Textsorten wurden Transkriptionen von akademischen Vorträgen und Interviews über berufliche und Alltagsthemen analysiert, dazu auch Dialoge zu unterschiedlichen Themen aus Internetforen, die sowohl Elemente der Schriftlichkeit als auch der Mündlichkeit enthalten und eine Sonderstellung im Vergleich mit herkömmlichen Kommunikationsformen einnehmen.

In Kapitel 8 werden zudem manuelle und semi-automatische Annotationsverfahren verglichen. Während andere Kohäsionsmitteln teilweise automatisch vorannotiert und nur noch manuell nachkorrigiert werden konnten, ist dies bei Ellipsen nicht möglich, da man hierfür nicht automatisch nach bestimmten Wörtern oder Oberflächenformen suchen kann, sondern nur nach potentiellen Ellipsenumgebungen. Automatische

Abfragen können lediglich dazu dienen, wenige Fälle zu finden und führen dabei stets zu einer Vielzahl an irrelevanten Treffern, die dann manuell wieder aussortiert werden müssen. Des Weiteren muss jeweils genau geprüft werden, ob es sich um eine kohäsive oder nicht-kohäsiv gebrauchte Ellipse oder eine andere Art von Fragment handelt und ob ein Antezedens im Text vorhanden ist. Eine manuelle Annotation kohäsiver Ellipsen sowie ihrer gegebenenfalls vorhandenen Antezedenzen erwies sich als bedeutend genauer als automatische Identifikationsmethoden. Die inzwischen im Korpus annotierten Kohäsionsmittel können nun gezielt abgefragt werden, und die jeweiligen Beispielkontexte und Häufigkeitswerte können extrahiert und anschließend statistisch ausgewertet werden.

In Kapitel 9 werden die Ergebnisse der Korpusstudie mit quantitativen Methoden ausgewertet. Zuerst werden die englischen und deutschen Daten der originalsprachlichen Texte ausgewertet. Dann werden Originale und Übersetzungen verglichen. Schließlich werden Diskrepanzen zwischen geschriebener und gesprochener Sprache und Registerunterschiede ausgewertet. Die Daten werden grafisch und tabellarisch aufgearbeitet und mithilfe von Methoden der deskriptiven Statistik analysiert und bewertet. Hierbei werden geeignete Kennzahlen berechnet und eventuelle Korrelationen zwischen verschiedenen Merkmalen sowie statistische Signifikanzen überprüft. Wie erwartet, zeigen die Ergebnisse, dass Ellipsen im Vergleich zu anderen Mitteln relativ selten verwendete Kohäsionsmittel sind. Überraschenderweise stellte sich heraus, dass die

englischen und deutschen Korpussektionen insgesamt fast identische Häufigkeiten an Ellipsen aufweisen. Ursprünglich erwarteten wir, im Englischen mehr kohäsive Ellipsen zu finden, was sich jedoch nicht bestätigte. In den deutschen Daten ließ sich allerdings eine größere Varianz beobachten. Mehr Texte als im Englischen beinhalten beispielsweise gar keine Ellipsen oder aber überdurchschnittlich viele. Es konnte festgestellt werden, dass die verschiedenen Arten von Ellipsen als Kohäsionsmittel besonders häufig in gesprochenen Texten in beiden Sprachen vorkommen.

Anschließend beschäftigt sich die Arbeit mit den Schlussfolgerungen der vorher ausgewerteten Untersuchung. Die Ergebnisse sind sowohl für die vergleichende Sprachforschung von Relevanz, aber auch für den Fremdsprachenunterricht und die Ausbildung von Übersetzern. Übersetzer, Sprachlerner und Autoren fremdsprachlicher Texte sollten sich sprachübergreifend über typische syntaktische Muster und Kohäsionsmittel in unterschiedlichen Kommunikationsszenarien bewusst sein. Wenn stilistisch markierte Konstruktionen wie Ellipsen von Sprachlernern falsch oder mit untypischen Häufigkeiten verwendet werden, können Text auf auffällige oder auch subtilere Weise Merkmale übersetzter oder fremdsprachlich produzierter Texte enthalten, stellenweise sogar redundant, zu explizit oder syntaktisch vereinfacht wirken. Durch Veränderungen in Bezug auf die Art der verwendeten Kohäsionsmittel in Übersetzungen können sich Verschiebungen semantischer Zusammenhänge oder ungünstige Distanzen zwischen anaphorischen

Ausdrücken und Antezedenzen im Zieltext ergeben.

Mit dieser Studie soll eine diskursorientierte vergleichende Grammatik der englischen und deutschen Sprache angeregt werden. Es ist geplant, eine zusammenfassende, kontrastive Beschreibung aller Kohäsionsmittel für das Englische und Deutsch im Rahmen des GECCo-Projektes zu erstellen. Des Weiteren wurde im Rahmen dieser Arbeit eine umfangreiche Sammlung von authentischen Belegen an anaphorischen Ellipsen und Satzfragmenten in verschiedenen Textsorten angelegt, welche in Form von Übungsblättern mit Übersetzungs- und Analyseaufgaben im Fremdsprachenunterricht und in der Ausbildung von Sprachmittlern und Autoren multilingualer Texte Verwendung finden kann. Das GECCo-Korpus kann für Abfragen von annotierten Kohäsionsmitteln in der Lehre genutzt werden. Verschiedene übersetzerische Lösungen können anhand von authentischen Textbelegen verglichen werden, um sich wichtige Aspekte textueller Kohäsions- und Kohärenzstrukturen bewusst zu machen. Die Arbeit bietet also sowohl Anschlussmöglichkeiten für die Praxis als auch für die weitere Forschung.

References

- Abels, Klaus (2004). Right Node Raising: Ellipsis or Across the Board Movement? In: Proceedings of NELS 34, 45–60.
- Abney, Steve (1987). The English Noun Phrase in its Sentential Aspect. Cambridge, Mass.: MIT.
- Abraham, Werner (2001): Modal Verbs: Epistemics in German and in English In: Frits Beukema, Sjef Barbiers and Wim van der Wurff (eds). Modality and Its Interaction with the Verbal System. Amsterdam: Benjamins.
- Abraham, Werner (2010). Methodische Überlegungen zu Grammatikalisierung, zyklischem Wandel und dem Wechsel von Analytik zu Synthetik im Deutschen – und zyklisch weiter zu Analytik (?). In: Dagmar Bittner and Livio Gaeta (eds): Kodierungstechniken im Wandel. Das Zusammenspiel von Analytik und Synthese im Gegenwartsdeutschen. Berlin/New York: de Gruyter, 249–274.
- Aelbrecht, Lobke (2010). The Syntactic Licensing of Ellipsis. Amsterdam/New York: John Benjamins
- Ágel, Vilmos (1991). Lexikalische Ellipsen. Fragen und Vorschläge. In: Zeitschrift für germanistische Linguistik 19, 24–48.
- Ágel, Vilmos and Mathilde Hennig (eds) (2007). Zugänge zur Grammatik der gesprochenen Sprache. Tübingen: Niemeyer.
- Alexiadou, Artemis, Liliane Haegeman and Melita Stavrou (2007). Noun Phrase in the Generative Perspective. Berlin/New York: Mouton de Gruyter.
- Altmann, Hans (1981). Formen der ‘Herausstellung’ im Deutschen: Rechtsversetzung, Linksversetzung, Freies Thema und verwandte Konstruktionen. Tübingen: Niemeyer.
- Arregi, Karlos (2010). Ellipsis in Split Questions. Natural Language and Linguistic Theory 28.3: 539–592.
- Arregui, Ana, Charles Clifton Jr., Lyn Frazier and Keir Moulton (2006). Processing Elided Verb Phrases with Flawed Antecedents: The Recycling Hypothesis. Journal of Memory and Language 55: 232–246.
- Auer, Peter (1988). On Deixis and Displacement. In: Folia Linguistica 22, 263–292.
- Auer, Peter and Susanne Günthner (2005). Die Entstehung von Diskursmarkern im Deutschen – ein Fall von Grammatikalisierung?. In: Torsten Leuschner, Tanja Mortelmans and Sarah de Groot (eds): Grammatikalisierung im Deutschen (Linguistik – Impulse und Tendenzen; 9.) Berlin: de Gruyter, 335–362.
- Auer, Peter (2007). Syntax als Prozess. In: Heiko Hausendorf (eds). Gespräch als Prozess. Linguistische Aspekte der Zeitlichkeit verbaler Interaktion. Tübingen: Narr, 95–124.
- Baker, Mona (1992). In Other Words: A Coursebook on Translation. London: Routledge.
- Baker, Mona (1993). Corpus Linguistics and Translation Studies – Implications and Applications. In: Mona Baker, Francis Gill and Elena

- Tognini-Bonelli (eds). *Text and Technology: In Honour of John Sinclair*. Amsterdam and Philadelphia: John Benjamins: 233–250.
- Barker, Chris (1995). *Possessive Descriptions*. Stanford, Cal.: CSLI Publications.
- Barron, Anne (2003). *Acquisition in Interlanguage Pragmatics: Learning How to Do Things with Words in a Study Abroad Context*, John Benjamins, Amsterdam.
- Barton, Ellen (1990). *Nonsentential Constituents – A Theory of Grammatical Structure and Pragmatic Interpretation*. John Benjamins: Amsterdam.
- Bauer, Laurie, Rochelle Lieber and Ingo Plag (2013). *The Oxford Reference Guide to English Morphology*. Oxford: OUP.
- Bhatt, Christa (1990). *Die syntaktische Struktur der Nominalphrase im Deutschen*. Tübingen: Narr.
- Beaugrande, Robert de (1980). *Text, Discourse, and Process*. Norwood, NJ: Ablex.
- Beaugrande, Robert de (1991). *Linguistic Theory: the Discourse of Fundamental Works*. London: Longman.
- Beaugrande, Robert de and Wolfgang Ulrich Dressler (1981). *Einführung in die Textlinguistik*, Tübingen: Niemeyer.
- Becher, Viktor (2011). *Explicitation and Implication in Translation. A Corpus-based Study of English-German and German-English Translations of Business Texts*. Doctoral dissertation, Universität Hamburg.
- Beck, Sigrid and Remus Gergel (2014). *Contrasting English and German Grammar: An Introduction to Syntax and Semantics*. Berlin/New York: Mouton de Gruyter.
- Behrens, Heike (1993). *Early Encoding of Temporal Reference in German*. In: *Proceedings of the 24th Annual Child Language Research Forum [1992]* Stanford, CA: CSLI. 61-71.
- Berka, Karel (1999). *On Frege's Philosophy of Language – a Linguistic Approach*. *Organon F: filozofický časopis*, Vol. 6 (2): 111–118.
- Berlage, Eva (2014). *Noun Phrase Complexity in English (Studies in English Language)*. Cambridge: CUP.
- Betten, Anne (1976). *Ellipsen, Anakoluthe und Parenthesen. Fälle für Grammatik, Stilistik, Sprechakttheorie oder Konversationsanalyse?* In: *Deutsche Sprache* 3: 207–230.
- Biber, Douglas (1988). *Variation across Speech and Writing*, Cambridge: CUP.
- Biber, Douglas and Edward Finegan (1989). *Drift and Evolution of English Style: a History of Three Genres*, In: *Language* 65: 487–517.
- Biber, Douglas (1990). *Methodological Issues Regarding Corpus-based Analyses of Linguistic Variation*. In: *Literary and Linguistic Computing* 5/3: 257–269.
- Biber, Douglas and Edward Finegan (1992). *The Linguistic Evolution of Five Written and Speech-based English Genres from the 17th to the 20th Centuries*, In: Matti Rissanen et al. (eds). *History of Englishes: New Methods and Interpretations in Historical Linguistics*. Berlin: Mouton. 688-704.
- Biber, Douglas (1993). *Representativeness in Corpus Design*. In: *Literary and*

- Linguistic Computing 8/4: 243–57.
- Biber, Douglas and Edward Finegan (1997). Diachronic Relations among Speech-based and Written Registers in English. In: T. Nevalainen and L. Kahlas-Tarkka (eds). *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*, Helsinki: Société Neophilologique. 253–275. (Reprinted in Conrad and Biber (eds) (2001), 66–83.)
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Bisiada, Mario (2013). *From Hypotaxis to Parataxis: An Investigation of English–German Syntactic Convergence in Translation*. Doctoral dissertation. University of Manchester.
- Bloomfield, Leonard (1933). *Language*. London: Allen and Unwin.
- Blum-Kulka, Shoshana (1986). Shifts of Cohesion and Coherence in Translation. In: Juliane House and Shoshana Blum-Kulka (eds). *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunther Narr, 17–35.
- Bolinger, Dwight L. (1978). Asking more than One Thing at a Time. In: Henry Hiz (eds). *Questions*, Dordrecht: Reidel, 107–150.
- Bolinger, Dwight L. (1980). *Language: The Loaded Weapon*. London: Longman
- Bonitz, Petra-Kristin (2012). Auxiliares *tun* als Evidenz für eine IP im Deutschen – der deutsche do-support. In: *Linguistik im Nordwesten: Beiträge zum 3. Nordwestdeutschen Linguistischen Kolloquium, Oldenburg 2010*, Bochum: Universitätsverlag Dr. N. Brockmeyer, 1–28.
- Bos, Johan and Jennifer Spenader. (2011). An Annotated Corpus for the Analysis of VP Ellipsis. *Language Resources and Evaluation* 45(4): 463–494.
- Bošković, Željko (2004). Two Notes on Right Node Raising. *University of Connecticut Working Papers in Linguistics* 12: 13–24.
- Bouton, Lawrence (1970). Antecedent-contained Pro Forms. In: *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society* (Chicago, IL).
- Bowman, Elizabeth (1966). *The Minor and Fragmentary Sentences of a Corpus of Spoken English*, The Hague: Mouton et al.
- Brazil, David (1995). *A Grammar of Speech*. Oxford: OUP.
- Broekhuis, Hans and Marcel den Dikken (2012). *Syntax of Dutch. Nouns and Noun Phrases Vol. 2*. Amsterdam: Amsterdam University Press.
- Brown, Penelope and Stephen C. Levinson (1987). *Politeness: Some Universals in Language Usage*. Cambridge: CUP.
- Bubenheimer, Felix (2001). *Grammatische Besonderheiten gesprochener Sprache und didaktische Konsequenzen für den DaF-Unterricht*. Manuscript at: <http://www.deutschservice.de/felix/daf/gesprkom.html> [last checked: 15/02/2016]
- Bühler, Karl (1990 [1934]). [Sprachtheorie. Die Darstellungsfunktion der Sprache, Jena, Fischer] *Theory of Language. The Representational*

- Function of Language, Amsterdam: John Benjamins.)
- Burridge, Kate (1998). From Modal Auxiliary to Lexical Verb: The Curious Case of Pennsylvania German *wotte*. In: Historical Linguistics 1995. Selected Papers from the 12th ICHL, Vol. 2: 19–31.
- Busler, Christine und Peter Schlobinski (1997). ‘Was er [schon] [...] konstruieren kann – das sieht er [oft auch] als Ellipse an.’ Über ‚Ellipsen‘, syntaktische Formate und Wissensstrukturen. In: Peter Schlobinski (eds). Syntax des gesprochenen Deutsch. Opladen: Westdeutscher, 93-115.
- Buss, Mareike (2004). Die Ellipse – ein linguistischer Kategorienfehler?. In: Trans 15. Internet-Zeitschrift für Kulturwissenschaften.
- Bussmann, Hadumod (1996) (eds). Routledge Dictionary of Language and Linguistics. London: Routledge.
- Buyko Ekaterina, Katrin Tomanek and Udo Hahn (2007). Resolution of Coordination Ellipses in Biological Named Entities using Conditional Random Fields. In: Proceedings of PACLING 2007, 163-171.
- Carlson, Katy (2001). Parallelism and Prosody in the Processing of Ellipsis Sentences. New York: Routledge.
- Carnie, Andrew (2013). Syntax: A Generative Introduction. 3rd edition. Oxford: Wiley-Blackwell.
- Carter, Ronald and Michael McCarthy (1995). Grammar and Spoken Language. Applied Linguistics 16 (2): 2. 141–158
- Carter, Ronald, McCarthy, Michael (2006). Cambridge Grammar of English: a Comprehensive Guide; Spoken and Written English Grammar and Usage. Cambridge: CUP.
- Chae Jeongmin, Younhee Jung, Taemin Lee, Soonyoung Jung, Chan Huh, Gilhan Kim, Hyeoncheol Kim, Heungbum Oh (2014). Identifying Non-elliptical Entity Mentions in a Coordinated NP with ellipses. Journal of Biomedical Informatics 47, 139-152.
- Chao, Wynn (1987). On Ellipsis. Doctoral dissertation, University of Massachusetts Amherst. Published 1988 by Garland, New York.
- Cheng, Lisa Lai Shen (1997). On the typology of wh-questions. New York/London, Garland Publishing.
- Chesterman, Andrew (1997). Memes of Translation. The Spread of Ideas in Translation Theory. Amsterdam/Philadelphia: John Benjamins.
- Chesterman, Andrew (2004). Hypotheses about Translation Universals. In: Gyde Hansen, Kirsten Malmkjær and Daniel Gile (eds). Claims, changes and Challenges in Translation Studies: Selected contributions from the EST congress, Copenhagen 2001, Amsterdam: John Benjamins, 1–13.
- Chomsky, Noam (1955). Logical Structure of Linguistic Theory, manuscript (published in 1975, New York: Plenum Press).
- Chomsky, Noam (1965). Aspects of the Theory of Syntax. Cambridge: MIT Press.
- Clarke, Benjamin Peter (2012). Do Patterns of Ellipsis in Text Support Systemic Functional Linguistics’ ‘Context-metafunction Hook-up’ Hypothesis? A Corpus-based Approach. Doctoral dissertation, Cardiff University.
- Clematide, Simon (2009). Koordination im Deutschen und ihre syntaktische Desambiguierung. Doctoral dissertation, Zurich University.

- Collins, Peter (2008). The English Modals and Semi-modals: Regional and Stylistic Variation. In: *The Dynamics of Linguistic Variation*. Amsterdam: John Benjamins
- Collins, Peter (2012). Grammatical Variation in English Worldwide: the Role of Colloquialization. In: *Linguistics and the Human Sciences*. Vol. 8.3: 289–306.
- Corr, Andreas (2010). Fragmente zwischen Störung und Anschlussfähigkeit – Theoretische Untersuchungen zur Ellipse als Absenz-Phänomen. In: Anke Grutschus, Peter Krilles (eds). *Figuren der Absenz*. Berlin: Frank & Timme, 87–101.
- Corver, Norbert and Marjo van Koppen (2008). Let's Focus on NP-ellipsis. *GAGL (Groninger Arbeiten zur Germanistischen Linguistik)* 48: 3–26.
- Cowan, Ron (2008). *The Teacher's Grammar of English with Answers: A Course Book and Reference Guide*. Cambridge: CUP.
- Craenenbroeck, Jeroen van (2004). *Ellipsis in Dutch Dialects*. LOT Dissertation Series 96. LOT.
- Craenenbroeck, Jeroen van (2010). *The Syntax of Ellipsis: Evidence from Dutch Dialects*. New York, NY: OUP.
- Craenenbroeck, Jeroen van and Anikó Lipták (2006). The Cross-linguistic Syntax of Sluicing: Evidence from Hungarian Relatives. *Syntax. A Journal of Theoretical, Experimental and Interdisciplinary Research*, 9 (3), 248–274.
- Craenenbroeck, Jeroen van and Jason Merchant (2013). Ellipsis Phenomena. In: Marcel den Dikken, (eds). *The Cambridge Handbook of Generative Syntax*. Cambridge: CUP, 701–745.
- Crystal, David (2003). *The Cambridge Encyclopedia of the English Language*. Cambridge: CUP.
- Culicover, Peter and Ray Jackendoff (2005). *Simpler Syntax*. Oxford: OUP.
- Dahl, Östen (1973). On So-called Sloppy Identity. *Synthese* 26: 81–112.
- Dalrymple, Mary, Stuart M. Shieber and Fernando C. N. Pereira (1991). Ellipsis and Higher Order Unification. *Linguistics and Philosophy* 14: 399–452.
- Dalrymple, Mary (2004). Against Reconstruction in Ellipsis. In: Reinaldo Elugardo and Robert J. Stainton (eds). *Ellipsis and Nonsentential Speech*. Dordrecht: Kluwer, 31–55.
- Delisle, Jean, Hannelore Lee-Jahnke and Monique C. Cormier (1999). Terminologie de la Traduction. Translation Terminology. Terminología de la Traducción. Terminologie der Übersetzung. Amsterdam: John Benjamins Publishing.
- Deppermann, Arnulf (2006). Construction Grammar – Eine Grammatik für die Interaktion? In: Arnulf Deppermann, Reinhard Fiehler and Thomas Spranz-Fogasy (eds). *Grammatik und Interaktion. Untersuchungen zum Zusammenhang von grammatischen Strukturen und Gesprächsprozessen*. Radolfzell: Verlag für Gesprächsforschung, 43–65.
- Diewald, Gabriele (1993). Zur Grammatikalisierung der Modalverben im Deutschen. *Zeitschrift für Sprachwissenschaft* 12, 218–234.
- Diewald, Gabriele (1999). *Die Modalverben im Deutschen: Grammatikalisierung und Polyfunktionalität*. Tübingen: Niemeyer.

- den Dikken, Marcel (2005). Specificational Copular Sentences and Pseudoclefts. In: Martin Everaert Henk and Van Riemsdijk (eds). *The Blackwell Companion to Syntax*, Vol. IV, Chap. 61. Oxford: Blackwell
- Doherty, Monika (1999). Position and Explicitness – Language Specific Conditions for the Use of Adverbial Clauses in Translations between German and English. In: Monika Doherty (ed.): *Sprachspezifische Aspekte der Informationsverteilung*. Berlin: Akademie-Verlag. 112–148.
- Doherty, Monika (2006). *Structural Propensities. Translating Nominal Word Groups from English into German*. (Benjamins Translation Library 65) Amsterdam/Philadelphia: Benjamins.
- Doleschal, Ursula and Edgar Hoffmann (2004). *Marktwirtschaftlicher Diskurs im Russischen*. In: Robert Hammel and Ljudmilla Geist (eds). *Linguistische Beiträge zur Slavistik*. Munich: Otto Sagner Verlag.
- Downing, Angela (2014). *English Grammar: A University Course*. 3rd edition. London/New York: Routledge.
- Duczmal, Stanislaw (1985). Ellipsis as a Cohesive Device in English and Polish: a Contrastive Analysis. In: *Poznan Studies in Contemporary Linguistics*, Vol. 20, 115-128.
- Dudenredaktion (eds). Peter Eisenberg, Jörg Peters, Peter Gallmann, Cathrine Fabricius-Hansen, Damaris Nübling, Irmhild Barz, Thomas A. Fritz, Reinhard Fiehler (2009). *Der Duden in zwölf Bänden*. Vol. 4. *Duden – Die Grammatik – Unentbehrlich für richtiges Deutsch*. 8th edition. Mannheim [et al.]: Dudenverlag.
- Duffield, Nigel and Ayumi Matsuo (2009). Native-Speakers' vs. L2 learners' sensitivity to parallelism in VP-ellipsis, *Studies in Second Language Acquisition* 31, 1-31.
- Dürscheid, Christa and Jan Georg Schneider (eds) (2015). *Handbuch Satz, Äußerung, Schema*. Berlin/Boston: de Gruyter (= HSW 4).
- Eguren, Luis (2009). Adjectives and Deleted Nominals in Spanish. In: Enoch Aboh, Elisabeth van der Linden, Josep Quer and Petra Sleeman (eds). *Romance Languages and Linguistic Theory*. Amsterdam/Philadelphia: John Benjamins, 67–86.
- Emmott, Catherine (1997). *Narrative Comprehension: a Discourse Perspective*. Oxford: OUP.
- Emonds, Joseph (2007). *Discovering Syntax: Clause Structures of English, German and Romance*. Berlin/New York: Mouton de Gruyter.
- Erben, Johannes (1969). 'Tun' als Hilfsverb im heutigen Deutsch, In: Ulrich Engel, Paul Grebe and Heinz Rupp (eds). *Festschrift für Hugo Moser zum 60. Geburtstag am 19. Juni 1969*. Düsseldorf: Pädagogischer Verlag Schwann, 46–52.
- Evert, Stefan (2005). *The CQP Query Language Tutorial*. IMS: Universität Stuttgart.
- Fabricius-Hansen, Cathrine (2010). *Deutsch im Kontrast – textbezogen*. In: Antje Dammel, Sebastian Kürschner and Damaris Nübling (eds). *Kontrastive Germanistische Linguistik*. Vol. 1. Hildesheim et al., 171–200.
- Fairclough, Norman (1994). *Conversationalization of Public Discourse and the Authority of the Consumer*. In: Russel Keat, Nigel Whiteley and Nicholas Abercrombie (eds). *The Authority of the Consumer*: London:

- Routledge, 253–268.
- Feilke, Helmuth, Klaus-Peter Kappert and Clemens Knobloch (2001). Grammatikalisierung, Spracherwerb und Schriftlichkeit – Zur Einführung ins Thema. In: Helmuth Feilke, Klaus-Peter Kappert and Clemens Knobloch (eds): *Grammatikalisierung, Spracherwerb und Schriftlichkeit*. Tübingen: Niemeyer, 1–28.
- Fernández, Raquel and Jonathan Ginzburg (2002). Non-sentential Utterances in Dialogue: A Corpus-based Study. In: *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, ACL Special Interest Group on Dialog, Philadelphia, USA, 15–26.
- Ferrara, Kathleen (1992). The Interactive Achievement of a Sentence: Joint Productions in Therapeutic Discourse. In: *Discourse Processes* 15, 207–228.
- Féry, Caroline and Katharina Hartmann (2001). Focus and Prosodic Structures of German Gapping and Right Node Raising. *Linguistics in Potsdam* 16: 79–115.
- Feist, James Murray (2008). *The Order of Premodifiers in English Nominal Phrases*. Doctoral dissertation, University of Auckland.
- Feist, Jim (2011). *Premodifiers in English: Their Structure and Significance*. Cambridge: CUP.
- Fiehler, Reinhard, Barden, Birgit, Elstermann, Mechthild and Barbara Kraft (2004). *Eigenschaften gesprochener Sprache*. Tübingen: Narr.
- Fiengo, Robert and Robert May (1994). *Indices and Identity*. Cambridge, Mass: MIT Press.
- Finegan, Edward (1998). English Grammar and Usage. In: Suzanne Romaine (ed.) *The Cambridge History of the English Language*. Vol. 4, 1776-1997, Cambridge: CUP, 536–588.
- Fischer, Klaus (1997). *German-English Verb Valency. A contrastive analysis*. Tübingen: Narr.
- Fischer, Klaus (2013). *Satzstrukturen im Deutschen und Englischen. Typologie und Textrealisierung*. Berlin: Akademie Verlag.
- Flowerdew, John (2012). *Discourse in English Language Education*. Abingdon, UK: Routledge.
- Forsgren, Kjell-Åke (1992). Zum Problem des Satzbegriffs im Deutschen. In: *Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB)* 114, 1, 3–27.
- Frazier Lyn and Charles Clifton Jr (1998). Comprehension of Sluiced Sentences. *Language and Cognitive Processes* 13: 499–520.
- Frazier, Lyn and Charles Clifton (2000). On Bound Variable Interpretations: The LF-only Hypothesis, *Journal of Psycholinguistic Research* 29: 125–139.
- Frazier Lyn and Charles Clifton Jr. (2001). Parsing Coordinates and Ellipsis: Copy. *Syntax*. 2001; 4(1):1–22.
- Frazier, Lyn and Charles Clifton Jr. (2005). The Syntax-discourse Divide: Processing Ellipsis. *Syntax*, 8(2): 121-174.
- Frazier, Lyn and Charles Clifton Jr. (2006). Ellipsis and Discourse Coherence. *Linguistics & Philosophy*, 29 (3): 315–346.
- Frazier, Lyn (2013). A Recycling Approach to Processing Ellipsis. In: Lisa

- Cheng and Norbert Corver (eds) *Diagnosis in Syntax*. Oxford: OUP.
- Frazier, Michael (2015). *Morphological Recoverability in Gapping*. Doctoral dissertation, Northwestern University, Evanston, Illinois.
- Gabelentz, Hans Georg Conon von der (1874-75). Weiteres zur vergleichenden Syntax. In: *Zeitschrift für Völkerpsychologie und Sprachwissenschaft* 8: 129–165.
- Garnham, Alan and Jane Oakhill (1987). Interpreting Elliptical Verb Phrases. *Quarterly Journal of Experimental Psychology*, 324, 611–627.
- Gengel, Kirsten (2013). *Pseudogapping and Ellipsis*. Oxford: OUP.
- Gergel, Remus (2005). *Modality and Ellipsis: Diachronic and Synchronic Evidence*. Doctoral dissertation, University of Tübingen.
- Giannakidou, Anastasia and Stavrou, Melita (1999). Nominalization and Ellipsis in the Greek DP. *The Linguistic Review* 16 (4): 295–331.
- Ginneken, Jacobus van (1910). Ellipsomanie [Ellipsomania]. *De Nieuwe Taalgids*, 4, 192-211.
http://www.dbnl.org/tekst/_taa008191001_01/_taa008191001_01_0031.php
 [last checked 01/02/2016]
- Givón, Talmy (1979). *On Understanding Grammar*. New York: Academic Press.
- Goldberg, Lotus (2005). *Verb-Stranding VP Ellipsis: A Cross-linguistic study*. Doctoral dissertation, McGill University.
- Goodwin, Charles (1995). Sentence Construction within Interaction. In: Quasthoff, Uta (ed.): *Aspects of Oral Communication*. Berlin: de Gruyter, 198–219.
- Graffi, Giorgio (2001). *200 Years of Syntax. A Critical Survey*. *Studies in the History of the Language Sciences* 98. Amsterdam: Benjamins.
- Grawe, Klaus (1991). Über den Umgang mit Zahlen. In: Klaus Grawe, Rolf Hänni, Norbert Semmer, Franziska Tschan (eds). *Über die richtige Art, Psychologie zu betreiben*. Göttingen: Hogrefe. 89–105.
- Greenbaum, Sidney, and Gerald Nelson (1999). Elliptical Clauses in Spoken and Written English. In: Peter Collins and Lee David (eds). *The Clause in English: in Honour of Rodney Huddleston*. Amsterdam: Benjamins.
- Grice, Herbert Paul (1975). Logic and Conversation. In: Peter Cole and Jerry L. Morgan (eds). *Syntax and Semantics 3: Speech Acts*, New York: Academic Press. 41–58.
- Gries, Stefan Th. (2009). *Quantitative Corpus Linguistics with R: A Practical Introduction*. London/New York: Routledge, Taylor & Francis Group.
- Gries, Stefan Th. (2013). *Statistics for Linguistics with R: A Practical Introduction* (2nd, revised edition). Berlin/Boston, MA: de Gruyter.
- Gunter Richard (1963). Elliptical Sentences in American English. In: *Lingua* 12, 137-150.
- Günther, Christine (2011). Noun Ellipsis in English: adjectival modifiers and the role of context. In: *Language and Linguistics*, Vol. 15, Special Issue 02: 279–301.
- Günther, Christine (2013). *The Elliptical Noun Phrase in English: Structure and Use*. New York: Routledge.
- Günther, Christine (forthcoming). The Rich, the Poor, the Obvious – Arguing for an Ellipsis Approach to “Adjectives used as Nouns”. In: Alex Chung-

- Hang Leung and Wim van der Wurff (eds). NPs in English: past and present. Amsterdam: John Benjamins.
http://www.anglistik.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische_Fakultaet/Anglistik_und_Amerikanistik/English_Language_and_Linguistics/Dateien/Detailseiten/Guenther/Guenther_2015_The_rich_the_poor_the_obvious.pdf [last checked 22/02/2015]
- Günthner, Susanne (1999). Entwickelt sich der Konzessivkonnektor obwohl zum Diskurs-marker? Grammatikalisierungstendenzen im Gesprochenen Deutsch. In: *Linguistische Berichte* 180, 409–446.
- Günthner, Susanne (2013). Ko-Konstruktionen im Gespräch: Zwischen Kollaboration und Konfrontation GIDI-Arbeitspapier 49 (08/2013), Münster University, <http://noam.uni-muenster.de/gidi/arbeitspapiere/arbeitspapier49.pdf> [last checked 10/03/2016]
- Gy, Pierre (1992). *Sampling of Heterogeneous and Dynamic Material Systems Theories of Heterogeneity, Sampling, and Homogenizing*. Amsterdam: Elsevier.
- Habermas, Jürgen (1962). *Strukturwandel der Öffentlichkeit – Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*. Neuwied: Hermann Luchterhand Verlag.
- Haegeman, Liliane (1990). Understood Subjects in English Diaries. *Multilingua*, 9(2):157–199.
- Haider, Hubert (1988). Die Struktur der deutschen Nominalphrase. In: *Zeitschrift für Sprachwissenschaft* 7/1988: 32–59.
- Haider, Hubert (1992). Die Struktur der Nominalphrase. Lexikalische und funktionale Strukturen. In: *Jahrbuch 1991 des Instituts für deutsche Sprache*. Mannheim, 304–333.
- Hall, Alison (2009). Subsentential utterances, ellipsis, and pragmatic enrichment. *Pragmatics & Cognition* 17:2, 222–250.
- Hall, Christopher (2010). Kontrastive Analyse Englisch-Deutsch. In: Hans-Jürgen Krumm, Christian Fandrych, Britta Hufeisen, Claudia Riemer (eds): *Deutsch als Fremd- und Zweitsprache*. Vol. 1. Berlin et al. de Gruyter, 550–561.
- Halliday M. A. K. (1967). Notes on transitivity and theme in English Part 2. *Journal of Linguistics* 3, 199–244.
- Halliday, M.A.K and Ruqaiya Hasan (1976). *Cohesion in English*. London: Longman.
- Halliday, M.A.K (1985). *Introduction to Functional Grammar*. London: Edward Arnold.
- Halliday, M.A.K (1994). *Introduction to Functional Grammar* (2nd edition). London: Edward Arnold.
- Halliday, M.A.K and Christian Matthiessen (2004). *An Introduction to Functional Grammar* (3rd edition). London: Arnold.
- Halliday, M.A.K. and Christian Matthiessen (2013). *Halliday's Introduction to Functional Grammar*. Fourth Edition. London: Routledge.
- Hankamer, Jorge (1971). *Constraints on Deletion in Syntax*, Doctoral dissertation, University of Yale, New Haven, CT published in 1979 under the title *Deletion in Coordinate Structures* [Outstanding Dissertations in

- Linguistics 16]. New York: Garland.
- Hankamer, Jorge (1973). Unacceptable Ambiguity, In: *Linguistic Inquiry* 4, 17–68.
- Hankamer, Jorge and Ivan Sag (1976). Deep and Surface Anaphora. In: *Linguistic Inquiry* 7:391–426.
- Hankamer, Jorge (1978). On the Nontransformational Derivation of Some Null VP Anaphors. In: *Linguistic Inquiry* 9: 66–74.
- Hankamer, Jorge and Ivan Sag (1984). Toward a theory of anaphoric processing. In: *Linguistics and Philosophy* 7: 325–345.
- Hansen-Schirra Silvia (2011). Between normalization and shining-through: Specific properties of English–German translations and their influence on the target language. In: Silvia Kranich, Viktor Becher et al. (eds) *Multilingual Discourse Production: Diachronic and Synchronic Perspectives*. Amsterdam: John Benjamins, 133–162.
- Hansen-Schirra, Silvia, Stella Neumann and Erich Steiner (2012). *Cross-linguistic Corpora for the Study of Translations. Insights from the language pair English – German*. Series: Text, Translation, Computational Processing. Berlin/New York: Mouton de Gruyter.
- Harbusch, Karin and Gerard Kempen (2007). Clausal coordinate ellipsis in German: The TIGER treebank as a source of evidence. In: *Proceedings of the Sixteenth Nordic Conference of Computational Linguistics (NODALIDA 2007)*. Tartu, Estonia.
- Harbusch, Karin (2011). Incremental Sentence Production and Clausal Coordinate Ellipsis: A Treebank Study Comparing Spoken and Written Language in Dutch and German, *Dialogue and Discourse* 5 (2011), 313–332.
- Hardt, Daniel (1992a). VP Ellipsis and Semantic Identity. *Proceedings of the Second Conference on Semantics and Linguistic Theory*. Edited by Chris Barker and David Dowty. Columbus, OH.
- Hardt, Daniel (1992b). An Algorithm for VP Ellipsis. In: *Proceedings of the 30th Annual Meeting*, Newark, DE. Association for Computational Linguistics.
- Hardt, Daniel (1993). *Verb Phrase Ellipsis: Form, Meaning, and Processing*. Ph.D. Dissertation. University of Pennsylvania, Distributed as IRCS Report 93-23.
- Hardt, Daniel (1997). VPE as a Proform: Some Consequences for Binding *Proceedings of CSSP 97*.
- Hardt, Daniel and Owen Rambow (2001). Generation of VP ellipsis: A corpus-based approach. In: *ACL 01: Proceedings of the 39th annual meeting on association for computational linguistics*, 290–297.
- Hartmann, Dietrich (1979). Syntaktische Eigenschaften und Funktionen von Satzpartikeln im Deutschen. In: Weydt, H. (ed.) *Die Partikeln der deutschen Sprache*. Berlin/New York: de Gruyter, 121–138.
- Hartmann, Katharina (2000). Right Node Raising and Gapping. *Interface Conditions on Prosodic Deletion*. Amsterdam: John Benjamins.
- Haspelmath, Martin (2007). Coordination. In: Shopen, Timothy (ed.) *Language typology and syntactic description, Vol. II: Complex constructions*. 2nd edition. Cambridge: CUP, 1–51.

- Haspelmath, Martin (2010a). Comparative Concepts and Descriptive categories in cross-linguistic studies. *Language* 86(4).
- Haspelmath, Martin (2010b). The interplay between comparative concepts and descriptive categories (reply to Newmeyer). *Language* 86: 696–699.
- Hawkins, John A. (1986). *A Comparative Typology of English and German: Uni-fying the Contrasts*. London and Sydney: Croom Helm.
- Hawkins, John A. (1988). The unity of English/German contrasts: Inferring a typology parameter. In: Caroline Duncan-Rose and Theo Vennemann (eds) *On Language, Rhetorica, Phonologica, Syntactica: A Festschrift for Robert P. Stockwell from his Friends and Colleagues*. London and New York: Routledge, 361–380.
- Hawkins, John A. (2014). *Cross-Linguistic Variation and Efficiency*. Oxford: OUP.
- Heim, Irene (1985). Notes on Comparatives and Related Matters. Manuscript, Austin: University of Texas at Austin. <http://semantics.uchicago.edu/kennedy/classes/f07/comparatives/heim85.pdf> [21/07/2016]
- Helasvuo, Marja-Liisa (2004). Shared Syntax: the Grammar of Co-constructions. In: *Journal of Pragmatics* 36: 1315–1336.
- Helbig, Gerhard and Joachim Buscha (1994). *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Leipzig: Langenscheidt.
- Hennig, Mathilde (2006). *Grammatik der gesprochenen Sprache in Theorie und Praxis*. Kassel: Kassel University Press.
- Hennig, Mathilde (2010). Elliptische Junktion in der Syntax des Neuhochdeutschen. In: Hans Ulrich Schmid (eds). *Perspektiven der germanistischen Sprachgeschichtsforschung (Jahrbuch für germanistische Sprachgeschichte 1)*, Berlin/New York, 76–103.
- Hill, Archibald A. (1958). *Introduction to Linguistic Structures: from Sound to Sentence in English*. New York: Harcourt Brace.
- Hinds, John (1982). *Ellipsis in Japanese*. Carbondale, Alberta: Linguistic Research.
- Hirschbühler, Paul (1982). VP-Deletion and Across-the Board Quantifier Scope. In: *Proceedings of NELS 12*, 132-139.
- Hjulmand, Lise-Lotte and Helge Schwarz (2012). *A Concise Contrastive Grammar of English*. 4th revised and expanded edition. Samfundslitteratur: Copenhagen.
- Hobbs, Jerry R., and Andrew Kehler (1997). A Theory of Parallelism and the Case of VP Ellipsis. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, Madrid, Spain*: 394-40.
- Hoeksema, Jack (2006). Pseudogapping: its syntactic analysis and cumulative effects on its acceptability. In: *Research on language and computation* 4, 335–352.
- Hoffmann, Christian (2012). *Cohesive Profiling: Meaning and Interaction in Personal Weblogs*. Amsterdam/Philadelphia: John Benjamins.
- Hoffmann, Ludger (1999). Ellipse und Analepse. In: Angelika Redder and Jochen Rehbein (eds). *Grammatik und mentale Prozesse*. Tübingen: Stauffenburg. (Stauffenburg Linguistik 7), 69-90.
- Hofmann, Eberhardt (2011). *Verhaltens- und Kommunikationsstile: Erkennen*

- und optimieren. Göttingen: Hogrefe.
- Höhle, Tilman (1983). Subjektlücken in Koordinationen. Ms., Univ. Tübingen. <http://www.ling.uni-potsdam.de/~thiersch/Hoehle83.pdf> [last checked 18/02/2016]
- Höhle, Tilman (1991). Assumptions about asymmetric coordination in German. In: Mascaró, Joan and Nespó, Marina (eds). *Grammar in progress*. Dordrecht: Foris, 221-235.
- Holmberg, Anders (2016). *The Syntax of Yes and No*. Oxford: OUP.
- House, Juliane (1989). Politeness in English and German: the functions of please and *bitte*. In: *Cross-cultural Pragmatics: Requests and Apologies*: 96–119.
- House, Juliane (2004). Linguistic Aspects of the Translation of Children's Books, in *Übersetzung-Translation-Traduction. An International Handbook*. Berlin/New York: Mouton de Gruyter.
- House, Juliane (2011). Linking Constructions in English and German Translated and Original Texts. In: Svenja Kranich, Viktor Becher, Steffen Höder and Juliane House (eds). *Multilingual Discourse Production: Diachronic and Synchronic Perspectives*. Amsterdam: John Benjamins, 163–182.
- Houser, Michael J. (2010a). *The Syntax and Semantics of 'do so' Anaphora*. Doctoral dissertation, University of California, Berkeley.
- Houser, Michael J. (2010b). *On the Anaphoric Status of Do So* http://linguistics.berkeley.edu/~mhouser/Papers/do_so_status [last checked 26/05/2016]
- Hopper, Paul J., and Elizabeth Closs Traugott. (1993). *Grammaticalization*. Cambridge: CUP.
- Huddleston, Rodney (1984). *Introduction to the Grammar of English*, Cambridge: CUP.
- Huddleston, Rodney (1988). *English Grammar: An Outline*, CUP.
- Huddleston Rodney and Geoffrey K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Hundt, Marianne and Christian Mair (1999). "Agile" and "uptight" genres: The corpus-based approach to language change in progress. In: *International Journal of Corpus Linguistics* 4, 221–242.
- Imo, Wolfgang (2011). Cognitions are not observable – but their consequences are: Mögliche Aposiopese-Konstruktionen in der gesprochenen Alltagssprache. In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 12, 265–300.
- İnce, Atakan (2009). *Dimensions of Ellipsis: Investigations in Turkish*. Doctoral dissertation, University of Maryland.
- Jackendoff, Ray (1971). Gapping and Related Rules. In: *Linguistic inquiry*, 21-35.
- Jackendoff, Ray (1977). *X-bar Syntax. A Study of Phrase Structure*. Linguistic Inquiry Monograph 2, MIT Press, Cambridge, Massachusetts.
- Jäger, Verena (submitted). Sprachwandel im 20. Jahrhundert. Ausdrücke deontischer Modalität im heutigen Englisch und Deutsch. In: *Studien zum deutsch-französischen Sprach- und Diskursvergleich*. Meidenbauer Verlag. [<http://www.english-linguistics.uni->

- mainz.de/Dateien/preprint_Veraenderungen_im_Gebrauch_von_MV.pdf
last checked 23/07/2016]
- Jeffries, Lesley (2006). *Discovering Language: the Structure of Modern English*. New York: Palgrave Macmillan.
- Jespersen, Otto (1933). *Essentials of English Grammar*. London: Allen & Unwin.
- Jespersen, Otto (1937). *Analytic Syntax*, London: Allen & Unwin.
- Johnson, Frances C. (1994). A Classification of Ellipsis Based on a Corpus of Information Seeking Dialogues. *Information Processing & Management*, 30(3), 315-325.
- Johnson, Kyle (1996). When Verb Phrases Go Missing. *Glott International* 2:5, 3-9.
- Johnson, Kyle (2001). What VP Ellipsis can do, and what it can't, but not why. Mark Baltin and Chris Collins (eds). *The handbook of contemporary syntactic theory*. Blackwell. 439-479.
- Jonkers, Sara and Ruigendijk, Esther (2013). Modal Verbs and their Relation to Finiteness in Dutch First Language Acquisition. In: Stavroula Stavrakaki (ed.). *Advances in Language Acquisition*, 111–121.
- Jung, Young Hee, Kinam Park , Jeong Min Chae, Soon Young Jung (2015). Resolution of Ellipses in Coordinate Noun Phrases, *Indian Journal of Science and Technology*, Vol 8(21): 1–5.
- Jürgens, Frank (1999). *Auf dem Weg zu einer pragmatischen Syntax. Eine vergleichende Fallstudie zu Präferenzen in gesprochen und geschrieben realisierten Textsorten*. Tübingen: Niemeyer. [Reihe Germanistische Linguistik. 2007]
- Juzwa, Urszula (2006). *The Syntax of Ellipsis in English and Polish: A Comparative View*. Doctoral dissertation, Adam Mickiewicz University.
- Kayne, Richard (2015). English 'One' and 'Ones' as Complex Determiners, published online at <http://linguistics.as.nyu.edu/docs/IO/2652/Kayne0615One.pdf> [last checked 23/02/2016]
- Kehler, Andrew (1993). A Discourse Copying Algorithm for Ellipsis and Anaphora Resolution. In: *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics (EACL-93)*, Utrecht, The Netherlands.
- Kehler, Andrew (1995). *Interpreting Cohesive Forms in the Context of Discourse Inference*. Ph.D. dissertation, Harvard University.
- Kehler, Andrew (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23: 533–575.
- Kehler, Andrew (2002). *Coherence in Discourse*. CSLI. Publications, Stanford, Calif.
- Keizer, Evelien (2007). *The English Noun Phrase. The Nature of Linguistic Categorization*. Cambridge: CUP.
- Kertz, Laura (2010a). *Ellipsis Reconsidered*. Doctoral dissertation, University of California, San Diego.
- Kertz, Laura (2010b). Verb Phrase Ellipsis: The View from Information Structure. *Language* 89(3):390–428.
- Kester, Ellen-Petra and Sleeman, Petra (2002). N-ellipsis in Spanish. In: Hans

- Broekhuis and Paula Fikkert (eds). *Linguistics in the Netherlands 2002*, Amsterdam/Philadelphia: John Benjamins, 107–116.
- Kim, Christina S. and Jeffrey T. Runner (2009). *Strict Identity, Coherence, and Parallelism in VP Ellipsis*. In: *Proceedings of Semantics and Linguistic Theory 19*, Columbus, OH: CLC Publishing.
- Kim, Christina, Gregory Kobele, Jeffrey Runner and John Hale (2011). *The Acceptability Cline in VP Ellipsis*. *Syntax* 14(4), 318–354.
- Klein, Katarina (2009). *Semi-modal Variation*. In: Andreas Dufter/Jürg Fleischer/Guido Seiler (eds): *Describing and Modeling Variation in Grammar*. Berlin: Mouton de Gruyter.
- Klein, Wolfgang (1981). *Some Rules of Regular Ellipsis in German*. In: Wolfgang Klein and Willem Levelt (eds). *Crossing the Boundaries in Linguistics. Studies Presented to Manfred Bierwisch*, Dordrecht: Reidel, 51–78.
- Klein, Wolfgang (1984). *Bühler Ellipse*. In: C.F. Graumann and Th. Herrmann (eds). *Karl Bühlers Axiomatik: Fünfzig Jahre Axiomatik der Sprachwissenschaft (117-141)*. Frankfurt am Main: Klostermann.
- Klein, Wolfgang (1985). *Ellipse, Fokusgliederung und thematischer Stand*. In: Reinhard Meyer-Herrmann and Hannes Rieser (eds). *Ellipsen und fragmentarische Ausdrücke Vol.1.31*. Tübingen: Niemeyer, 1–24.
- Klein, Wolfgang (1993). *Ellipse*. In: Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann (eds). *Syntax: Ein internationales Handbuch zeitgenössischer Forschung*, Berlin: de Gruyter, 763–798.
- Klein, Wolfgang and Vater, Heinz (1998). *The Perfect in English and German*. In: Leonid Kulikov and Heinz Vater (eds). *Typology of Verbal Categories. Papers presented to Vladimir Nedjalkov on the occasion of his 70th birthday* Tübingen: Niemeyer, 215–235.
- Koch, Peter and Wulf Oesterreicher (1985). *Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte*. *Romanistisches Jahrbuch* 36/85: 15-43.
- Kolk, Herman and Claus Heeschen (1992). *Agrammatism, paragrammatism and the management of language*. *Language and Cognitive Processes*, 7, 82-129.
- Kolk, Herman and Ben Hofstede (1994). *The Choice For ellipsis: A case study of stylistic shifts in an agrammatic speaker*. *Brain and Language*, 47, 507-509.
- Kolk, Herman (2001). *Does Agrammatic Speech Consitute a Regression to Child Language? A three-way Comparison between Agrammatic, Child and Normal Ellipsis*. *Brain and Language*, 77, 340–350.
- König, Ekkehard (1982). *Scalar Particles in German and their English Equivalents*. In: Walter F. W. Lohnes and Edwin A. Hopkins (eds). *Ann Arbor: Karoma Publishers*.
- König, Ekkehard (2001). *Kontrastive Analysen Deutsch Englisch: eine Übersicht*. In: Gerhard Helbig, Lutz Götze, Gert Henrici and Hans-Jürgen Krumm (eds). *Deutsch als Fremdsprache. Ein internationales Handbuch*. Berlin: de Gruyter, 324–330.

- König, Eckehard and Volker Gast (2012). *Understanding English-German Contrasts. Grundlagen der Anglistik und Amerikanistik*. Berlin: Erich Schmidt Verlag. [3rd, extended edition].
- König, Katharina (2012). Reconstructing the point of reference in the stand-alone deswegen-micro-aposiopesis, *GIDI-Arbeitspapier 42* (07/2012), Münster University, <http://noam.uni-muenster.de/gidi/arbeitspapiere/arbeitspapier42.pdf> [last checked 07/02/2016]
- Königs, Karin (2011). *Übersetzen Englisch – Deutsch. Lernen mit System*, 3rd edition, Munich: Oldenbourg Verlag.
- Kortmann, Bernd and Paul Georg Meyer (1992). Is English grammar more explicit than German grammar after all? In: Christian Mair and Manfred Markus (eds). *New Departures in Contrastive Linguistics – Neue Ansätze in der kontrastiven Linguistik*. Innsbruck: Universität Innsbruck, 155-166.
- Kortmann, Bernd (2001). In the Year 2525... Reflections on the Future Shape of English. In: *Anglistik 12*: 97–114.
- Kramer, Ruth and Kyle Rawlins (2009). Polarity Particles: an Ellipsis Account. In: *Proceedings of NELS 39*.
- Krifka, Manfred (2001). For a Structured Meaning Account of Questions and Answers. In: Caroline Fery and Wolfgang Sternfeld (eds). *Audiatu Vox Sapientia. A Festschrift for Arnim von Stechow*, Akademie Verlag, 287–319.
- Krisch, Thomas (2009). On the ‘Syntax of Silence’ in Proto-Indo-European. In: Roland Hinterhölzl and Svetlana Petrova (eds). *Information Structure and Language Change. New Approaches to Word Order Variation*. Berlin/New York: Mouton de Gruyter, 191–220.
- Krug, Manfred G. (1998). Gotta – The tenth Central Modal in English? Social, Stylistic and Regional Variation in the British National Corpus as Evidence of Ongoing Grammaticalization. In: *The Major Varieties of English (Papers from MAVEN 97)*: 177–191.
- Krug, Manfred G. (2000). *Emerging English Modals: A Corpus-based Study of Grammaticalization*. Berlin & New York: Mouton de Gruyter.
- Kufner, Herbert L. (1962). *The Grammatical Structures of English and German*. Chicago: Chicago University Press.
- Kuno, Susumo (1976). Gapping: A Functional Analysis. In: *Linguistic Inquiry 7*, 300–318.
- Kunz, Kerstin (2010). *English and German Nominal Co-reference. A Study of Political Essays*. Frankfurt am Main: Peter Lang.
- Kunz, Kerstin and Erich Steiner (2013). Cohesive Substitution in English and German: a Contrastive and Corpus-based Perspective. In: Karin Aijmer and Bengt Altenberg (eds). *Advances in Corpus-based Contrastive Linguistics. Studies in honour of Stig Johansson*. Amsterdam: John Benjamins. 201-231.
- Kunz, Kerstin and Ekaterina Lapshinova-Koltunski (2015). Cross-linguistic Analysis of Discourse Variation Across Registers. In: Karin Aijmer and Hilde Hasselgård (eds). *Cross-linguistic Studies at the Interface between Lexis and Grammar*. *Nordic Journal of English Studies*. Vol. 14 (1), 258-288.

- Kunz, Kerstin, Degaetano-Ortlieb, Stefania, Lapshinova-Koltunski, Ekaterina, Menzel, Katrin and Steiner, Erich. (forthcoming a). English-German Contrasts in Cohesion and Implications for Translation. In: Gert de Sutter, Isabelle Delaere, and Marie-Aude Lefer (eds). *Empirical Translation Studies. New Theoretical and Methodological Traditions*. TILSM series. Mouton de Gruyter.
- Kunz, Kerstin, Ekaterina Lapshinova-Koltunski, José Manuel Martínez Martínez, Katrin Menzel, Erich Steiner (forthcoming b). Shallow Features as Indicators of English-German Contrasts in Lexical Cohesion. In: *Languages in Contrast*.
- Labov, William (1969). Contraction, Deletion, and Inherent Variability of the English Copula. *Language* 45: 715–62.
- LaCara, Nicholas (2010). Verbal Ellipsis in the Nominal Domain, Master's Thesis, University of California Santa Cruz, people.umass.edu/nlacara/papers/vend.pdf [last checked 10.03.2016]
- Lakoff, Robin (1972). The pragmatics of modality, In: *Papers from the 8th regional meeting*. Chicago Linguistic Society, 229–246.
- Lange de, Joke (2008). Article Omission in Headlines and Child Language: A Processing Approach. *LOT Dissertation Series* 189. LOT.
- Langer, Nils (2001). *Linguistic Purism in Action. How auxiliary 'tun' was Stigmatized in Early New High German*. Berlin/New York: de Gruyter.
- Lapshinova-Koltunski, Ekaterina, Kerstin Kunz and Marilisa Amoia (2012). Compiling a Multilingual Spoken Corpus. In: Heliana Mello, Massimo Pettorino, Tommaso Raso (eds). *Proceedings of the VIIth GSCP International Conference: Speech and Corpora*, Firenze: Firenze University Press, 79–84.
- Lappin, Shalom (1992). The Syntactic Basis of Ellipsis Resolution. *Proceedings of the Stuttgart Ellipsis Workshop*, 1-47.
- Lappin, Shalom and Hsue-Hueh Shih (1996). A Generalized Reconstruction Algorithm for Ellipsis Resolution in *Proceedings of COLING-96*, 687–692.
- Lasnik, Howard (1999a). Verbal Morphology: Syntactic Structures Meets the Minimalist Program. In: Howard Lasnik (eds). *Minimalist Analysis*. Blackwell, Malden, Mass. 97–119.
- Lasnik, Howard (1999b). Pseudogapping Puzzles. In: Shalom Lappin and Elabbas Benmamoun (eds). *Fragments. Studies in Ellipsis and Gapping*. Oxford: OUP, 141-174.
- Lasnik, Howard (2010). On Ellipsis: Is Material that is Phonetically Absent but Semantically Present Present or Absent Syntactically? In: Hans Götzche (ed.). *Memory, Mind and Language*. Cambridge Scholars Publishing, 221–242.
- Lasnik, Howard (2013). Multiple Sluicing in English? In: *Syntax* 17(1), 1–20.
- Lechner, Winfried (2004). *Ellipsis in Comparatives*. Berlin/New York: Mouton de Gruyter (*Studies in Generative Grammar* 72).
- Lee, David Y. W. (2001). Genres, Registers, Text types, Domains and Styles: Clarifying the Concepts and Navigating a Path through the BNC Jungle. *Language Learning and Technology*, Vol.5 (3): 37–72.
- Leech, Geoffrey (2003). *Modality on the Move: The English Modal*

- Auxiliaries 1961–1992. In: Roberta Facchinetti, Manfred Krug and Frank Palmer (eds). *Modality in Contemporary English*. Berlin/New York: Mouton de Gruyter: 223–240.
- Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith (2009). *Change in Contemporary English. A Grammatical Study*. Cambridge: CUP.
- Leech Geoffrey (2013). Where have all the Modals gone? An Essay on the Declining Frequency of Core Modal Auxiliaries in Recent Standard English. In: Marín-Arrese Juana I., Marta Carretero Lapeyre, Jorge Arús Hita and Johan van der Auwera (eds). *English Modality: Core, Periphery and Evidentiality*. Berlin: Mouton de Gruyter: 95–115.
- Lehrndorfer, Anne (2013). *Didaktik des übersetzungsgerechten Schreibens*. In: Baumann, Klaus-Dieter and Hartwig Kalverkämpe (eds): *Theorie und Praxis des Dolmetschens und Übersetzens in fachlichen Kontexten*, Berlin: Frank & Timme, 657–678.
- Leisi, Ernst and Christian Mair (2008). *Das heutige Englisch: Wesenszüge und Probleme*. Heidelberg: Universitätsverlag Winter (9th edition).
- Leonard, Sterling Andrus (1929). *The Doctrine of Correctness in English Usage, 1700-1800*. Madison: Univ. of Wisconsin Press.
- Lerner, Gene H. (1987). *Collaborative Turn Sequences: Sentence Construction and Social Action*. Doctoral dissertation. University of California at Irvine.
- Levin, Nancy (1978). Some Identity-of-sense Deletions Puzzle me. Do they you? In: *Proceedings of the 14th Annual Meeting of the Chicago Linguistic Society*. Chicago University, 229–240.
- Levin Nancy (1979). *Main Verb Ellipsis in Spoken English*. Doctoral dissertation, Ohio State University.
- Lightfoot, David (1999). *The Development of Language: Acquisition, Change, and Evolution*. Malden, Mass.: Blackwell.
- Linell, Per (2005). *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. London/New York: Routledge.
- Llombart-Huesca, Amalia. (2002). Anaphoric One and NP-Ellipsis. *Studia Linguistica* 56(1):59–89.
- Lobeck, Anne (1995). *Ellipsis: Functional Heads, Licensing and Identification*. New York: OUP.
- Lobin, Henning (1993). *Koordinationssyntax als prozedurales Phänomen*. Tübingen: Narr.
- Lugebil K. (1884). Zur Frage über zweitheilige und einheitliche Sätze. *Archiv für slavische Philologie* 8, 36–68
- Mair, Christian (1997). Parallel Corpora: A Real-Time Approach to the Study of Language Change in Progress. In: *Corpus-Based Studies in English: Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17)*, 195–208.
- Mair, Christian (2006). *Twentieth-Century English: History, Variation and Standardization*, Cambridge: CUP.
- Mair, Christian and Geoffrey Leech (2006). Current Changes in English Syntax. In: Bas Aarts and April McMahon (eds). *The Handbook of English Linguistics*. Oxford: Blackwell: 318–342.

- Makino, Seiichi (1993). Shooryaku no nichiei hikaku [Contrasting Japanese and English ellipsis]. *Nihongogaku* 12: 41–49.
- Marcus, Inge Hannelore Elfriede (1982). The Cohesive Relation of Verbal Ellipsis in English and its Counterparts in Portuguese, Master thesis, Federal University of Paraná, Brazil.
- Markus, Manfred and Josef Wallmannsberger (eds) (1987). *English-German Contrastive Linguistics: A Bibliography*. Frankfurt a.M.: Lang.
- Martin, Andrea Eyleen and Brian McElree (2008). A Content-addressable Pointer Mechanism Underlies Comprehension of Verb-phrase ellipsis. *Journal of Memory and Language* 58, 879–906.
- Martin, Andrea Eyleen and Brian McElree (2009). Memory Operations that Support Language Comprehension: Evidence from Verb-phrase ellipsis. *Journal of Experimental Psychology: Learning Memory and Cognition*, 35, 1231-1239.
- Martin, Andrea Eyleen and Brian McElree (2011). Direct-access Retrieval during Sentence Comprehension: Evidence from Sluicing. *Journal of Memory and Language*, 64, 327-343.
- Martin, J. R. and Rose, David (2003). *Working with Discourse: Meaning beyond the Clause*. London/New York: Continuum.
- Matthiessen, Christian, Kazuhiro Teruya and Marvin Lam (2010). Key terms in Systemic Functional Linguistics. Continuum International Publishing Group.
- Mayr, Ernst (1982). *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*, Cambridge, MA: Harvard University Press.
- McEnery, Tony (2003). Corpus Linguistics. In: Ruslan Mitkov (eds). *Oxford Handbook of Computational Linguistics*. Oxford: OUP. 448-463.
- McPherson, Glen (1990). *Statistics in Scientific Investigation: its basis, application*, New York: Springer.
- McShane, Marjorie J. (2005). *A Theory of Ellipsis*. OUP.
- McWhorter, John (2008). Why does a Language Undress? Strange Cases in Indonesia. In: Miestamo, Matti, Kaius Sinnemäki, Fred Karlsson (eds): *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins Publishing, 167–190.
- Menzel, Katrin (2016a). The Elusive Ellipsis – the Complex History of a Vague Grammatical Concept in Need of Empirical Grounding. In: *Lege artis. Language yesterday, today, tomorrow, The Journal of University of SS Cyril and Methodius in Trnava*. Warsaw: De Gruyter Open, Vol. 1 (1), June 2016, 163–201.
- Menzel, Katrin (2016b). Annotating Cohesive Ellipses in an English-German corpus. In: Sheena Gardner, Sian Alsop (eds). *Systemic Functional Linguistics in the Digital Age* (eds). *Equinox*, 310–327.
- Menzel, Katrin (2016c). Textkompetenz in der Fremdsprachenvermittlung und Übersetzer Ausbildung – ein korpusbasierter Sprach- und Registervergleich zu Kohäsionsmitteln im Englischen und Deutschen. In: Thomas Tinnefeld et al. (ed.). *Saarbrücker Schriften zu Linguistik und Fremdsprachendidaktik*. Saarbrücken: HTW, 133-152.
- Menzel, Katrin (2014a). Ellipsen als Stil- und Kohäsionsmittel in deutschen und englischen politischen Reden. In: *Germanistische Mitteilungen*.

- Zeitschrift für Deutsche Sprache, Literatur und Kultur, 40.1: Deutsch kontrastiv, 31–50.
- Menzel, Katrin (2014b). The Conceptualisation and Annotation of Ellipses in English and German Corpus Texts (GECCo Corpus) with MMAX2, Project Working Report (WP 2.1/2.4)
- Menzel, Katrin (2013). A Corpus Linguistic Study of Ellipsis as a Cohesive Device, In: Proceedings of Corpus Linguistics 2013, St. Petersburg, Russia, 86-93, <http://corpora.phil.spbu.ru/Works2013/Menzel.pdf>
- Menzel, Katrin and Ekaterina Lapshinova-Koltunski (2014). Kontrastive Analyse deutscher und englischer Kohäsionsmittel in verschiedenen Diskurstypen, In: *tekst i dyskurs – Text und Diskurs. Zeitschrift der Abteilung für germanistische Sprachwissenschaft des Germanistischen Instituts Warschau.* 247–266.
- Menzel, Katrin, Lapshinova-Koltunski, Ekaterina and Kerstin Kunz (forthcoming). Cohesion and Coherence in Multilingual Contexts. In: Katrin Menzel, Ekaterina Lapshinova-Koltunski and Kerstin Kunz (eds). *New Perspectives on Cohesion and Coherence: Implications for Translations (Series: Translation and Multilingual Natural Language Processing 6)*. Berlin: Language Science Press, 1–11.
- Merchant, Jason (2001). *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*. Oxford: OUP.
- Merchant, Jason (2004). Fragments and ellipsis. *Linguistics and Philosophy* 27: 661-738.
- Merchant, Jason (2008). An Asymmetry in Voice Mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry*, 39(1):169–179
- Merchant, Jason (2010a). Three Kinds of Ellipsis. In: Francois Recanati, Isidora Stojanovic, Neftali Villanueva (eds). *Context-Dependence, Perspective, and Relativity*, Berlin: de Gruyter. 141–192.
- Merchant, Jason (2010a). Voice Mismatches and the Dark Side of Ellipsis. Paper presented at the 2010 Syntax Fest, Indiana University. <http://www.indiana.edu/~lingdept/SyntaxFest/iu.3.voice.pronouns.NPIs.pdf> [15/05/2016]
- Merchant Jason and Andrew Simpson (eds) (2012). *Sluicing: Cross-linguistic perspectives*, Oxford: OUP.
- Merchant, Jason (2013). Voice and Ellipsis. *Linguistic Inquiry* 44:77–108.
- Merchant, Jason, Lyn Frazier, Charles Clifton, Jr. and Thomas Weskott (2013). Fragment Answers to Questions: A Case of Inaudible Syntax. In: Laurence Goldstein (eds). *Brevity*. Oxford: OUP, 21-35.
- Merchant, Jason (2016). Ellipsis: A Survey of Analytical Approaches. In: Jeroen van Craenenbroeck and Tanja Temmerman (eds). *Handbook of ellipsis*, OUP: Oxford.
- Miller, Philip (1990). Pseudogapping and *do so* Substitution. In: Proceedings of CLS 26, Vol. 1, Chicago Linguistic Society, Chicago, 293-305.
- Miller, Philip, and Geoffrey K. Pullum (2013). Exophoric VP Ellipsis. In: Philip Hofmeister and Elisabeth Norcliffe (eds). *The Core and the Periphery: Data-Driven Perspectives on Syntax Inspired by Ivan A. Sag*, 5-32. Stanford, CA: CSLI Publications.
- Miller, Philip (2014). A Corpus Study of Pseudogapping and its Theoretical

- Consequences. In: *Empirical Issues in Syntax and Semantics*, Vol. 10, 73–90.
- Miller, Philip (n.a.). Verb Phrase Ellipsis with Nominal Antecedents http://didel.script.univ-paris-diderot.fr/claroline/backends/download.php?url=L0FydGljbGVzX2FuZl9ib29rcy9NaWxsZXIuSGVtZm9ydGguVlBFLk5vbS5BbnQuc3VibWl0dGVkLnBkZg%3D%3D&cidReset=true&cidReq=41LNA154_001 [01.12.2014]
- Mönnink, Inge de (2000). *On the Move: The Mobility of Constituents in the English Noun Phrase – a Multi-Method Approach*. Amsterdam: Rodopi.
- Morgan, Jerry L. (1973). Sentence fragments and the notion ‘sentence’, in *Issues in Linguistics: Essays in honour of Henry and Rene Kahane*, UIP, Urbana.
- Müller Christoph and Strube, Michael (2006). Multi-Level Annotation of Linguistic Data with MMAX2'. In: Braun, Sabine, Kohn, Kurt and Mukherjee, Joybrato (eds). *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods*. Frankfurt: Peter Lang, 197–214.
- Murguia, Elixabete (2004). *Syntactic Identity and Locality Restrictions on Verbal Ellipsis*. Doctoral dissertation, University of Maryland.
- Nariyama, Shigeko (2004). Subject Ellipsis in English. *Journal of Pragmatics* 36: 237–264.
- Neumann, Stella (2013). *Contrastive Register Variation. A Quantitative Approach to the Comparison of English and German*. Trends in Linguistics. Studies and Monographs. Berlin/Boston: de Gruyter Mouton.
- Newmeyer, Frederick J. and Laurel B. Preston (eds) (2014). *Measuring grammatical complexity*. Oxford: OUP.
- Nielsen, Leif A. (2003). Verb Phrase Ellipsis Detection Using Machine Learning Techniques. *RANLP 2003*, 317–326
- Nielsen, Leif A. (2004). Verb Phrase Ellipsis Detection using Automatically Parsed Text. *COLING 2004*.
- Nielsen, Leif A. (2005). *A Corpus-based Study of Verb Phrase Ellipsis Identification and Resolution*. Doctoral dissertation, King’s College London.
- Nishigauchi, Taisuke (1998). ‘Multiple Sluicing’ in Japanese and the functional nature of wh-phrases. *Journal of East Asian Linguistics* 7:121–152.
- Ntelitheos, Dimitris (2004). *Syntax of Elliptical and Discontinuous Nominals*. Doctoral dissertation, UCLA.
- Nunes, Jairo and Cynthia Zocca (2009). Lack of Morphological Identity and Ellipsis Resolution in Brazilian Portuguese. In: Jairo Nunes (ed.). *Minimalist Essays on Brazilian Portuguese Syntax*, 215–236.
- Olsen, Susan (1989). Das Possessivum: Pronomen, Determinans oder Adjektiv? In: *Linguistische Berichte* 120 (1989): 133–153.
- Ong, Walter J. (1982). *Orality and Literacy: The Technologizing of the Word*. London & New York: Methuen.
- Ortner, Hanspeter. (1985). Welche Rolle spielen die Begriffe ‘Ellipse’, ‘Tilgung’, ‘Einsparung’ usw. in der Sprachbeschreibung? In: Reinhard

- Meyer-Hermann and Hannes Rieser, (eds): *Ellipsen und fragmentarische Ausdrücke II*. Tübingen: Niemeyer. (Linguistische Arbeiten. 148/2.), 165–202.
- Ortner, Hanspeter (1987). *Die Ellipse. Ein Problem der Sprachtheorie und der Grammatikschreibung*. Tübingen: Niemeyer.
- Otsuki, Kyoko (2009). *Cross-linguistic Study of Elliptical Utterances in Task-oriented Dialogues with Classroom Implications*. Doctoral dissertation, University of Edinburgh.
- Ott Dennis (2012). *Local Instability: Split Topicalization and Quantifier Float in German*. Berlin/New York: De Gruyter.
- Paul, Hermann (1995 [1880]). *Prinzipien der Sprachgeschichte*. 10th ed. Tübingen: Niemeyer.
- Payne, John (1993). The Headedness of Noun Phrases. Slaying the Nominal Hydra. In: Greville G. Corbett, Norman M. Fraser and Scott McGlashan (eds). *Heads in Grammatical Theory*. CUP, 114–139.
- Payne, John, Geoffrey K. Pullum, Barbara Scholz and Eva Berlage (2013). Anaphoric one and its implications. In: *Language* 89(4): 794–829.
- Penn, C., and Cleary, J. (1988). Compensatory Strategies in the Language of Closed Head Injured Patients. *Brain Injury*, 1, 3-17.
- Peters, Francis E. (1967). *Greek Philosophical Terms: A Historical Lexicon*. New York: New York University Press.
- Peters, Pam (2004). *The Cambridge Guide to English Usage*. Cambridge: CUP.
- Petrov, Slav, Dipanjan Das and Ryan McDonald (2012). A Universal Part-of-Speech Tagset. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA).
- Phillips, Colin and Dan Parker (2014). The Psycholinguistics of Ellipsis. *Lingua* 151, 78–95.
- Poirier, Josée, Katie Wolfinger, Lisa Spellman, Lisa and Lewis P. Shapiro (2010). The Real-Time Processing of Sluiced Sentences. *Journal of Psycholinguistic Research*. 39: 411–427.
- Polenz, Peter von (2008). *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens*, 3rd edition. Berlin/New York: de Gruyter.
- Portner, Paul and Raffaella Zanuttini (2005). The Semantics of Nominal Exclamatives. In: Reinaldo Elugardo and Robert J. Stainton (eds). *Ellipsis and Nonsentential Speech*, Kluwer Academic Publishers, 57-67.
- Postal, Paul (1974). *On Raising*. Cambridge: MIT Press.
- Powers, Willow R. (2005). *Transcription Techniques for the Spoken Word*. Lanham, MD: Altamira Press.
- Progovac, Ljiljana (2006). The Syntax of Nonsententials: Small Clauses and Phrases at the Root. In: Ljiljana Progovac, Kate Paesani, Eugenia Casielles, and Ellen Barton (eds). *The Syntax of Nonsententials: Multidisciplinary Perspectives*. Amsterdam: Benjamins.
- Provost, Gary (1988). *Beyond Style: Mastering the Finer Points of Writing*. Cincinnati, Ohio: Writer's Digest Books.
- Pullum, Geoffrey K. (2000). Hankamer Was! In: Miller Sandy Chung, Jim McCloskey and Nathan Sanders (eds). *Jorge Hankamer WebFest*:

- <http://babel.ucsc.edu/Jorge/> [last checked 10/03/2016]
- Quirk, Randolph and Greenbaum, Sidney and Leech, Geoffrey and Svartvik, Jan (1972). *A Grammar of Contemporary English*. London: Longman.
- Quirk, Randolph, and Sidney Greenbaum (1973a). *A University grammar of English*. London: Longman.
- Quirk, Randolph, and Sidney Greenbaum (1973b). *A Concise Grammar of Contemporary English*. New York: Harcourt.
- Quirk, Randolph, Sidney Greenbaum, Jan Svartvik, Geoffrey Leech (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Quirk, Randolph and Sidney Greenbaum (1990). *A Student's Grammar of the English Language*, Harlow: Longman, Ltd.
- Radford, Andrew (1993). Head-hunting. On the Trail of the Nominal Janus. In: Greville G. Corbett, Norman M. Fraser and Scott McGlashan (eds). *Heads in Grammatical Theory*. CUP, 73–113.
- Raible, Wolfgang (ed.). (1998). *Medienwechsel. Erträge aus zwölf Jahren Forschung zum Thema ‚Mündlichkeit und Schriftlichkeit‘*. Tübingen: Narr.
- Raible, Wolfgang (2001). Foundations: Theoretical Foundations of Language Universals and Language. In: Martin Haspelmath, Ekkehard König and Wulf Oesterreicher (eds). *Language Typology and Language Universals*. Berlin: Mouton de Gruyter, 1-39.
- Rath, Rainer (1979). *Kommunikationspraxis. Analysen zur Textbildung und Textgliederung im gesprochenen Deutsch*. Göttingen: Vandenhoeck und Ruprecht.
- Rath, Rainer (1985). Gesprochene und geschriebene Form der heutigen Standardsprache. In: Werner Besch, Oskar Reichmann and Stefan Sonderegger (eds). *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. Vol. 2, Berlin/New York: de Gruyter, 1651-1663
- Reich, Sabine (2001). *Struktur und Erwerb der englischen Nominalphrase*. Linguistische Arbeiten 437. Tübingen: Niemeyer.
- Reich, Ingo (2002). Question/Answer Congruence and the Semantics of wh-Phrases. In: *Theoretical Linguistics* 28, 73–94.
- Reich, Ingo (2003). *Frage, Antwort und Fokus*. Band 55 der Reihe *Studia Grammatica*. Akademie-Verlag: Berlin.
- Reich, Ingo (2007). Toward a Uniform Analysis of Short Answers and Gapping. In: Kerstin Schwabe and Susanne Winkler (eds). *On Information Structure, Meaning and Form*. Amsterdam: John Benjamins. 467-484.
- Reich, Ingo (2011). Ellipsis. In: Claudia Maienborn, Klaus von Stechow and Paul Portner (eds): *Semantics: An International Handbook of Natural Language Meaning*. Berlin/New York: de Gruyter. (Handbooks of Linguistics and Communication Science): 1849-1874.
- Reis, Marga (2001). Bilden Modalverben im Deutschen eine syntaktische Klasse? In: Müller, Reimar and Marga Reis (eds). *Modalität und Modalverben im Deutschen*. Hamburg: Buske. 287-318
- Repp, Sophie (2009a). Koordination, Subordination und Ellipse. *Linguistische Berichte*, 16, 245–265.
- Repp, Sophie (2009b). *Negation in Gapping*. Oxford: OUP.

- Ricento, Thomas (1987). Clausal Ellipsis in Multi-party Conversation in English. *Journal of Pragmatics* 11: 751-775.
- Ries, John (1931). *Beiträge zur Grundlegung der Syntax. Teil 3: Was ist ein Satz?* Prag: Taussig & Taussig.
- Rooryck, Johan (1985). Gapping-Zeugma in French and English: A Non-Deletion Analysis. *Linguistic Analysis* 15, 187-229.
- Rooryck, Johan and Erik Schoorlemmer (2013). False Identity: Ellipsis, Mismatches, and Underspecification. Conference presentation at Identity in Ellipsis, LUCL Leiden. <http://media.leidenuniv.nl/legacy/rooryck,-schoorlemmer-final.pdf> [last checked 31/01/2016]
- Ross, John R. (1969). Guess who? In: R. Binnick, A. Davidson, G. Green and J. Morgan (eds). *Papers from the fifth regional meeting of the Chicago Linguistic Society. Chicago Linguistic Society.* 252–286
- Ross, John R. (1970). Gapping and the Order of Constituents. In: Manfred Bierwisch, Karl Erich Heidolph (eds). *Progress in Linguistics.* Den Haag: Mouton, 249-259.
- Ruiter, Marina (2008). *Speaking in Ellipses: The Effect of a Compensatory Style of Speech on Functional Communication in Chronic Agrammatism.* Doctoral dissertation, Radboud University Nijmegen.
- Ruiter, Marina, Herman Kolk and Toni Rietveld (2010). *Speaking in Ellipses: The Effect of a Compensatory Style of Speech on Functional Communication in Chronic Agrammatism.* *Neuropsychological Rehabilitation*, 20, 423-458.
- Sag, Ivan (1976). *Deletion and Logical Form.* Ph.D. Dissertation, MIT.
- Sanders, Gerald A. (1977). A Functional Typology of Elliptical Coordination. In: Fred R. Eckman (eds). *Current Themes in Linguistics: Bilingualism, Experimental Linguistics, and Language Typologies.* Washington, DC: Hemisphere, 241-270.
- Sandig, Barbara (2000). Zu einer Gesprächs-Grammatik. Prototypische elliptische Strukturen und ihre Funktionen in mündlichem Erzählen. In: *Zeitschrift für germanistische Linguistik* 28, 291–318.
- Schachter, Paul (1977). Does She or Doesn't She? *Linguistic Inquiry* 8: 763–767.
- Schachter, Paul (1985). Part-of-Speech Systems. In: Timothy Shopen (ed.). *Language Typology and Syntactic Description. Vol. 1: Clause Structure.* Cambridge: CUP. 3–61.
- Scheffler, Tatjana (2005). Ellipsis, Subsential Speech, and the Contextualism Discussion. Unpublished manuscript, www.ling.upenn.edu/~tatjana/papers/scheffler-ellipsis.pdf [last checked 30/04/2016].
- Schiefer, Erhard (1974). Zur Abgrenzung von Nominalsatz und Ellipse. In: *Zeitschrift für Vergleichende Sprachforschung*, Vol. 88, 199-217.
- Schlangen, David (2003). *A Coherence-based Approach to the Interpretation of Non-sentential Utterances in Dialogue.* Doctoral dissertation, School of Informatics, University of Edinburgh.
- Schlenck, Klaus, Claudia Schlenck and Luise Springer (1995). *Die Behandlung des schweren Agrammatismus - Reduzierte-Syntax-Therapie (REST).* Stuttgart: Thieme.

- Schlobinski, Peter (1997). Zur Analyse syntaktischer Strukturen in der gesprochenen Sprache. In: Peter Schlobinski (ed). *Syntax des gesprochenen Deutsch*. Opladen: Westdeutscher Verlag. 9-26.
- Schmolz, Helene (2015). Anaphora Resolution and Text Retrieval. *A Linguistic Analysis of Hypertexts*, Berlin et al.: de Gruyter.
- Schwabe, Kerstin (1994). *Syntax und Semantik situativer Ellipsen*, Studien zur deutschen Grammatik, Gunter Narr Verlag.
- Schwarz, Bernhard (1998). Reduced Conditionals in German: Event Quantification and Definiteness. *Natural Language Semantics* 6, 271–301.
- Schwarz, Christian (2004). Die tun-Periphrase im Deutschen. Master thesis, <https://www.freidok.uni-freiburg.de/fedora/objects/freidok:1759/datastreams/FILE1/content> [last checked 04/06/2016]
- Schwitalla, Johannes (2003). *Gesprochenes Deutsch. Eine Einführung*. Berlin: Erich Schmidt.
- Selting, Magret (1997). Sogenannte 'Ellipsen' als interaktiv relevante Konstruktionen? Ein neuer Versuch über Reichweite und Grenzen des Ellipsenbegriffs für die Analyse gesprochener Sprache in der konversationellen Interaktion. In: Schlobinski, P. (ed.): *Syntax des gesprochenen Deutsch*. Opladen: Westdeutscher Verlag, 117–155.
- Shahabi, Mitra and Jorge Baptista (2012). A Corpus-based Translation Study on English-Persian Verb Phrase Ellipsis. In: *ICAME Journal, Computers in English Linguistics*, No. 36, 95-112.
- Shieber, Stuart M., Mary Dalrymple and Fernando C. N. Pereira. (1996). Interactions of scope and ellipsis. *Linguistics and Philosophy*, 19(5):527–552.
- Shopen, Tim and Walerian Świczekowski (2006). Some Remarks on Ellipsis in Polish and English. In: *Poznan Studies in Contemporary Linguistics*, Vol. 20, 111–126.
- Schubert, Christoph (2008). *Englische Textlinguistik: eine Einführung* Berlin: Erich Schmidt.
- Siemund, Peter (2004). 'English'. In: Thorsten Roelcke (ed.). *Variationstypologie. Variation Typology. Ein sprachtypologisches Handbuch der europäischen Sprachen in Geschichte und Gegenwart. A Typological Handbook of European Languages Past and Present*. Berlin/New York: de Gruyter, 1–29.
- Skeppstedt, Maria (2013). Annotating Named Entities in Clinical text by Combining Pre-annotation and Active Learning. In: *Proceedings of the ACL student research workshop, Sofia, Bulgaria*, 74–80.
- Sleeman, Petra (1996). *Licensing Empty Nouns in French*. Den Haag: Holland Academic Graphics.
- Smith, Noah A. (2001). Ellipsis Happens, Deletion is how. *University of Maryland Working Papers in Linguistics*, vol. 11, 176–191.
- Smith, Robin Deirdre (1986). *A Syntactic Quicksand: Ellipsis in Seventeenth and Eighteenth-century English Grammars*. Doctoral dissertation, Leiden, Delft: Eburon.
- Söll, Ludwig (1974). *Gesprochenes und geschriebenes Französisch*. Berlin: Erich Schmidt.
- Sommerfeldt Karl-Ernst, Starke, Günter and Werner Hackel (2011).

- Einführung in die Grammatik der deutschen Gegenwartssprache. Berlin/New York: de Gruyter.
- Stainton, Robert (2004). The Pragmatics of Non-sentences. In: Laurence R. Horn and Gregory Ward (eds): *The Handbook of Pragmatics*. 266–287.
- Stainton, Robert (2006a). *Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language*. Oxford: OUP.
- Stange, Ulrike (2016). *A Corpus-based Study on Variation in Acquisition, Function and Usage*. Amsterdam/Philadelphia: John Benjamins.
- Stein, Stephan (2003). *Textgliederung: Einheitenbildung im geschriebenen und gesprochenen Deutsch – Theorie und Empirie*. Berlin/New York: de Gruyter (Studia Linguistica Germanica, Band 69).
- Steiner, Erich (2015). Contrastive Studies of Cohesion and their Impact on our Knowledge of Translation. In: Zhang, Meifang and Jeremy Munday (eds). *Discourse Analysis and Translation*. Special issue of TARGET. *International Journal of Translation Studies*. Amsterdam: John Benjamins.
- Straumann, Heinrich (1935). *Newspaper headlines: A study in linguistic method*. London: George Allen & Unwin Ltd.
- Streit, Michael and Hans-Ulrich Krieger (2004). Ellipsis Resolution by Controlled Default Unification for Multi-modal and Speech Dialog Systems, ACL 2004: Workshop on Reference Resolution and its Applications, Barcelona, Spain, Association for Computational Linguistics, 87-94.
- Stockwell, Robert P. (1977). Motivations for Exbraciation in Old English. In: C.N. Li (ed.). *Mechanisms of Syntactic Change*, London/Austin: University of Texas Press, 291–314.
- Stubbs, Michael (2001). Computer-assisted Text and Corpus Analysis: Lexical Cohesion and Communicative Competence. In: Deborah Schiffrin, Deborah Tannen, Heidi E. Hamilton (eds). *Handbook of Discourse Analysis*. Oxford: Blackwell. 304–320.
- Sundby, Bertil, Anne Kari Bjørge and Kari E. Haugland (1991). *A Dictionary of English Normative Grammar 1700-1800*. Amsterdam: Benjamins.
- Sutherland, Sean (2016). *A Beginner's Guide to Discourse Analysis*. London: Palgrave Macmillan.
- Swan, Michael (2005). *Practical English Usage*. 3rd edition. Oxford: OUP.
- Sweet, Henry (1875–76). Words, Logic, and Grammar. In: *Transactions of the Philological Society 1875-76*: 470–503.
- Taboada, Maite (2004). *Building Coherence and Cohesion: Task-Oriented Dialogue in English and Spanish*. Amsterdam/Philadelphia: John Benjamins.
- Takahashi, Shoichi (2004). Pseudogapping and Cyclic Linearization. In: *Proceedings of NELS 34*. Amherst, Mass.: GLSA.
- Teich, Elke (2003). *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Berlin: de Gruyter.
- Tesak, Jürgen and Jürgen Dittmann (1991). Syntaktische Strukturen und Ellipsen in deutschen Telegrammen. In: Elisabeth Feldbusch, Reiner Pogarell and Cornelia Weiß (eds). *Neue Fragen der Linguistik. Akten des 25. Linguistischen Kolloquiums, Paderborn 1990*. Vol. 2: Innovation und

- Anwendung. Tübingen. 245-251.
- Tesak, Jürgen and Chris Code (2008). *Milestones in the History of Aphasia: Theories and Protagonists*. Hove, UK: Psychology Press.
- Thomas, Andrew L. (1979). Ellipsis: The Interplay of Sentence Structure and Context. In: *Lingua* 47, 43–68.
- Thomas, Andrew L. (1987). The Use and Interpretation of Verbally Determinate Verb Group Ellipsis in English, In: *International Review of Applied Linguistics*, 25 (1), 1–14.
- Thoms, Gary (2013). Lexical Mismatches in Ellipsis and the Identity Condition. In: *NELS* 42, 559–572.
- Toury, Gideon (1995). *Descriptive Translation Studies and beyond*. Amsterdam: John Benjamins.
- Traugott, Elizabeth Closs (1995). The Role of the Development of Discourse Markers in a Theory of Grammaticalization. Paper presented at the 12th International Conference on Historical Linguistics. <http://web.stanford.edu/~traugott/papers/discourse.pdf> [last checked 23/07/2016]
- Trotta, Joe (2000). *Wh-clauses in English: Aspects of Theory and Description*. Amsterdam: Rodopi.
- Tucker, Gordon H. (2007). 'Sorry to muddy the waters' – Accounting for Speech Act Formulae and Formulaic Variation in a Systemic Functional Model of Language. In: Christopher S. Butler, Raquel Hidalgo Downing and Julia Lavid (eds). *Functional Perspectives on Grammar and Discourse: In honour of Angela Downing*. [Studies in Language Companion Series 85], 395–418.
- Tucker, Gordon H. (2014). Towards a Solution to Adjective-noun 'fused-head' constructions in a systemic functional grammar. In: Alison Duguid, Anna Marchi, Alan Partington and Charlotte Taylor (eds). *Gentle Obsession: Literature, Linguistics and Learning in Honour of John Morley* Rome: Artemide.
- Tutin, Agnès, François Trouilleux, Catherine Clouzot, Eric Gaussier, Annie Zaenen, Stéphanie Rayot and Georges Antoniadis (2000). Annotating a Large Corpus with Anaphoric Links. In: *Proceedings of the 3rd Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, Lancaster.
- Valois, Daniel, Royle, Phaedra and Ann Sutton (2009). L'ellipse du nom en français: le rôle de l'acquisition pour la théorie linguistique. *The Canadian Journal of Linguistics* 54 (2): 339–366.
- Van der Auwera, Johan, Dirk Noël and An Van Linden (2013). Had better, 'd better and better: Diachronic and Transatlantic variation.' In: J. I. Marín Arrese, M. Carretero, J. Arús Hita and J. van der Auwera (eds). *English Modality: Core, Periphery and Evidentiality*. Berlin: Mouton de Gruyter. 119–153.
- Wachtel, Martin (2004). *Grammatik und vieles mehr. Linguistische Grundlagen und Lernziele für den Deutschunterricht der Sekundarstufen*. Frankfurt/M. et al.: Lang.
- Walker, Marilyn A. (1996). Limited Attention and Discourse Structure. *Computational Linguistics*, 22(2), 255–264.
- Waltereit, Richard (2006). *Abtönung. Zur Pragmatik und historischen*

- Semantik von Modalpartikeln und ihren funktionalen Äquivalenten in romanischen Sprachen. Tübingen: Niemeyer.
- Warner, Anthony R. (1986). Ellipsis Conditions and the Status of the English Copula. *York Papers in Linguistics*, 12, 153–172.
- Wasow, Thomas (1972). *Anaphoric Relations in English*, Doctoral dissertation, Massachusetts Institute of Technology.
- Wasow, Thomas (1979). *Anaphora in Generative Grammar*, E. Ghent: Story-Scientia.
- Webber, Bonnie Lynn (1978). *A Formal Approach to Discourse Anaphora*, Doctoral dissertation, Harvard University.
- Wegener, Philipp (1885). *Untersuchungen über die Grundfragen des Sprachlebens*. Halle: Max Niemeyer.
- Weir, Andrew (2012). Left-edge Deletion in English and Subject Omission in Diaries. *English Language and Linguistics* 16(1). 105–29.
- Weir, Andrew (2014). *Fragments and Clausal ellipsis*. Doctoral dissertation, University of Massachusetts Amherst.
- Westney, Paul (1995). *Modals and Periphrastics in English: an Investigation into the Semantic Correspondence between Certain English Modal Verbs and their Periphrastic Equivalents*. Tübingen: Max Niemeyer Verlag.
- Wilson, Peter (2000). *Mind the Gap: Ellipsis and Stylistic Variation in Spoken and Written English*, Harlow: Longman.
- Williams, Edwin (1977). Discourse and Logical Form. *Linguistic Inquiry* 8(1), 101–139.
- Winkler, Susanne (2003). *Ellipsis at the Interfaces: An Information Structural Investigation of Sentence-Bound and Discourse-Bound Ellipsis in English*. Habilitationsschrift. Tübingen University.
- Winkler, Susanne (2005). *Ellipsis and Focus in Generative Grammar*. (Studies in Generative Grammar 81.) Series Editors: Harry van der Hulst, Henk van Riemsdijk and Jan Koster). Berlin/New York: de Gruyter.
- Winkler, Susanne (2006). Ellipsis at the Interfaces: An Information-structural Proposal. In: Valéria Molnár and Susanne Winkler (eds). *The Architecture of Focus*. Studies in Generative Grammar. Berlin/New York: Mouton de Gruyter, 401–428
- Winkler, Susanne (2013). Syntactic Diagnostics for Extraction of Focus from Ellipsis site. In: L. Cheng Lisa and Norbert Corver (eds). *Diagnosing Syntax*. Oxford: OUP.
- Wurmbrand, Susi (2006). Verb Clusters, Verb Raising, and Restructuring. In: Martin Everaert and Henk van Riemsdijk (eds). *The Blackwell Companion to Syntax*, Oxford: Blackwell. 229–343.
- Wyngaerd, G. Vanden (1998). *Gapping Constituents*. Ms. Catholic University of Brussels, <https://lirias.kuleuven.be/bitstream/123456789/408979/1/09HRPL%26L02.pdf> [last checked 08/02/2016]
- Xiang Ming, Julian Grove and Jason Merchant (2014). *Ellipsis Sites Induce Structural Priming Effects*, University of Chicago, unpublished manuscript http://home.uchicago.edu/~juliang/ellipsis_priming.pdf [last checked 23/07/2016]
- Yamamoto, Kazuhide, Eiichiro Sumita, Osamu Furuse and Hitoshi Iida

- (1997). Ellipsis Resolution in Dialogues via Decision-Tree Learning. In: Proceedings of Natural Language Processing Pacific-Rim Symposium (NLPRS'97), 423–428.
- Yamamoto, Kazuhide and Eiichiro Sumita (1998). Feasibility Study for Ellipsis Resolution in Dialogues by Machine-Learning Technique, In: Proceedings of COLING '98 17th international conference on Computational linguistics, Vol. 2, 1428–1435.
- Yvel, Divad (2001). You've Got Funny Mail. U.S.: AuthorHouse.
- Zagona, Karen (1988a). Proper Government of Antecedentless VP in English and Spanish. In: Natural Language and Linguistic Theory 6: 95-128.
- Zagona, Karen (1988b). Verb Phrase Syntax: A Parametric Study of English and Spanish, Studies. In: Natural Language and Linguistic Theory, Kluwer, Dordrecht.
- Zebrowska, Ewa (2005). Review of [Albrecht Plewnia (2003). Sätze, denen nichts fehlt. Eine dependenzgrammatische Untersuchung elliptischer Konstruktionen. Am Beispiel des mitteldeutschen Dialekts des Ermlands. Hildesheim/Zürich/New York: Olms (Germanistische Linguistik, Monographien, 11).], Zeitschrift für Dialektologie und Linguistik Vol. 72 (2).
- Zifonun, Gisela, Ludger Hoffmann and Bruno Strecker (1997). Grammatik der deutschen Sprache. Berlin, New York: de Gruyter.
- Ziliak, Stephen T. and Deirdre N. McCloskey (2008). The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice and Lives, University of Michigan Press.
- Zwicky, Arnold (1987). Suppressing the Z's. In: Journal of Linguistics 23, 133–148.

Annex 1: Frequencies of ellipses and fragments in GECCo per text

Ellipses and fragments in GECCo - English Originals (per text)

a. = absolute figures, n. = normalised figures (per 1000 tokens), coh. = cohesive, non-coh. = non-cohesive, all = nominal, verbal / clausal

TEXT	tokens	nominal coh. a.	nominal coh. n.	verbal / clausal cohesive a.	verbal / clausal cohesive n.	all cohesive a.	all cohesive n.	non-coh. nominal a.	non-coh. nominal n.	non-coh. verbal / clausal a.	non-coh. verbal / clausal n.	all non-coh. a.	all non-coh. n.	other fragments a.	other fragments n.
EO_ESSAY_001	2278	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	6	2.63
EO_ESSAY_002	653	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	3.06
EO_ESSAY_003	725	0	0.00	0	0.00	0	0.00	1	1.38	1	1.38	2	2.76	3	4.14
EO_ESSAY_004	1132	1	0.88	0	0.00	1	0.88	3	2.65	1	0.88	4	3.53	2	1.77
EO_ESSAY_005	1790	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	18	10.06
EO_ESSAY_006	1058	0	0.00	0	0.00	0	0.00	1	0.95	0	0.00	1	0.95	8	7.56
EO_ESSAY_007	1557	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	7	4.50
EO_ESSAY_008	1737	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	6	3.45
EO_ESSAY_009	882	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
EO_ESSAY_010	815	0	0.00	0	0.00	0	0.00	1	1.23	0	0.00	1	1.23	3	3.68
EO_ESSAY_011	1020	0	0.00	0	0.00	0	0.00	2	1.96	0	0.00	2	1.96	5	4.90
EO_ESSAY_012	1010	2	1.98	0	0.00	2	1.98	1	0.99	1	0.99	2	1.98	6	5.94

EO_ESSAY_013	849	1	1.18	0	0.00	1	1.18	1	1.18	0	0.00	1	1.18	8	9.42
EO_ESSAY_014	840	1	1.19	0	0.00	1	1.19	0	0.00	0	0.00	0	0.00	1	1.19
EO_ESSAY_015	2461	3	1.22	0	0.00	3	1.22	1	0.41	2	0.81	3	1.22	2	0.81
EO_ESSAY_016	1092	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	1.83
EO_ESSAY_017	1138	1	0.88	0	0.00	1	0.88	0	0.00	0	0.00	0	0.00	2	1.76
EO_ESSAY_018	1168	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	3	2.57
EO_ESSAY_019	1809	2	1.11	2	1.11	4	2.21	2	1.11	1	0.55	3	1.66	8	4.42
EO_ESSAY_020	1856	0	0.00	0	0.00	0	0.00	1	0.54	0	0.00	1	0.54	6	3.23
EO_ESSAY_021	1065	0	0.00	1	0.94	1	0.94	0	0.00	0	0.00	0	0.00	3	2.82
EO_ESSAY_022	1122	1	0.89	0	0.00	1	0.89	0	0.00	0	0.00	0	0.00	3	2.67
EO_ESSAY_023	1012	1	0.99	0	0.00	1	0.99	0	0.00	0	0.00	0	0.00	2	1.98
EO_ESSAY_024	900	1	1.11	0	0.00	1	1.11	0	0.00	0	0.00	0	0.00	1	1.11
EO_ESSAY_025	1029	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	1.94
EO_ESSAY_026	885	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	3	3.39
EO_ESSAY_027	845	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	4	4.73
EO_ESSAY_028	1132	0	0.00	1	0.88	1	0.88	0	0.00	0	0.00	0	0.00	11	9.72
EO_ESSAY_029	1138	0	0.00	0	0.00	0	0.00	1	0.88	0	0.00	1	0.88	6	5.27
EO_FICTION_001	3683	3	0.81	4	1.09	7	1.90	5	1.36	7	1.90	12	3.26	28	7.60
EO_FICTION_002	3548	20	5.64	1	0.28	21	5.92	3	0.85	5	1.41	8	2.25	10	2.82
EO_FICTION_003	3955	3	0.76	7	1.77	10	2.53	2	0.51	0	0.00	2	0.51	39	9.86
EO_FICTION_004	3743	0	0.00	2	0.53	2	0.53	3	0.80	2	0.53	5	1.34	33	8.82
EO_FICTION_005	3613	1	0.28	1	0.28	2	0.55	1	0.28	8	2.21	9	2.49	26	7.20

EO_FICTION_006	3540	2	0.56	0	0.00	2	0.56	1	0.28	1	0.28	2	0.56	4	1.13
EO_FICTION_007	3676	6	1.63	0	0.00	6	1.63	0	0.00	2	0.54	2	0.54	29	7.89
EO_FICTION_008	3862	3	0.78	8	2.07	11	2.85	2	0.52	4	1.04	6	1.55	22	5.70
EO_FICTION_009	3756	3	0.80	5	1.33	8	2.13	2	0.53	2	0.53	4	1.06	14	3.73
EO_FICTION_010	3620	3	0.83	2	0.55	5	1.38	4	1.10	1	0.28	5	1.38	3	0.83
EO_INSTR_001	3593	0	0.00	0	0.00	0	0.00	0	0.00	1	0.28	1	0.28	52	14.47
EO_INSTR_002	3116	1	0.32	0	0.00	1	0.32	0	0.00	1	0.32	1	0.32	24	7.70
EO_INSTR_003	3583	1	0.28	0	0.00	1	0.28	1	0.28	0	0.00	1	0.28	10	2.79
EO_INSTR_004	3615	0	0.00	1	0.28	1	0.28	0	0.00	0	0.00	0	0.00	36	9.96
EO_INSTR_005	3528	0	0.00	0	0.00	0	0.00	0	0.00	1	0.28	1	0.28	7	1.98
EO_INSTR_006	4043	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	22	5.44
EO_INSTR_007	3501	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	53	15.14
EO_INSTR_008	3706	1	0.27	0	0.00	1	0.27	0	0.00	0	0.00	0	0.00	33	8.90
EO_INSTR_009	3780	0	0.00	1	0.26	1	0.26	0	0.00	1	0.26	1	0.26	15	3.97
EO_INSTR_010	3702	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	104	28.09
EO_POPSCI_001	3330	1	0.30	0	0.00	1	0.30	1	0.30	0	0.00	1	0.30	11	3.30
EO_POPSCI_002	3043	2	0.66	0	0.00	2	0.66	1	0.33	2	0.66	3	0.99	5	1.64
EO_POPSCI_003	3577	0	0.00	0	0.00	0	0.00	1	0.28	0	0.00	1	0.28	14	3.91
EO_POPSCI_004	3956	0	0.00	1	0.25	1	0.25	0	0.00	1	0.25	1	0.25	9	2.28
EO_POPSCI_005	3400	4	1.18	1	0.29	5	1.47	3	0.88	4	1.18	7	2.06	8	2.35
EO_POPSCI_006	3535	0	0.00	1	0.28	1	0.28	1	0.28	0	0.00	1	0.28	6	1.70
EO_POPSCI_007	1483	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00

EO_POPSCI_008	3332	3	0.90	1	0.30	4	1.20	3	0.90	2	0.60	5	1.50	9	2.70
EO_POPSCI_009	3296	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	4	1.21
EO_POPSCI_010	2882	0	0.00	0	0.00	0	0.00	3	1.04	1	0.35	4	1.39	4	1.39
EO_POPSCI_011	3314	0	0.00	0	0.00	0	0.00	2	0.60	1	0.30	3	0.91	8	2.41
EO_SHARE_001	1843	2	1.09	0	0.00	2	1.09	0	0.00	0	0.00	0	0.00	5	2.71
EO_SHARE_002	2571	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	4	1.56
EO_SHARE_003	2754	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	7	2.54
EO_SHARE_004	4409	8	1.81	1	0.23	9	2.04	6	1.36	4	0.91	10	2.27	18	4.08
EO_SHARE_005	4621	1	0.22	0	0.00	1	0.22	2	0.43	0	0.00	2	0.43	15	3.25
EO_SHARE_006	2573	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	15	5.83
EO_SHARE_007	3215	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	4	1.24
EO_SHARE_008	1842	0	0.00	1	0.54	1	0.54	1	0.54	1	0.54	2	1.09	8	4.34
EO_SHARE_009	2300	0	0.00	1	0.43	1	0.43	0	0.00	0	0.00	0	0.00	14	6.09
EO_SHARE_010	2726	0	0.00	1	0.37	1	0.37	1	0.37	0	0.00	1	0.37	18	6.60
EO_SHARE_011	3550	0	0.00	0	0.00	0	0.00	1	0.28	1	0.28	2	0.56	13	3.66
EO_SHARE_012	2564	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	10	3.90
EO_SHARE_013	856	0	0.00	1	1.17	1	1.17	0	0.00	0	0.00	0	0.00	4	4.67
EO_SPEECH_001	2650	2	0.75	0	0.00	2	0.75	2	0.75	2	0.75	4	1.51	12	4.53
EO_SPEECH_002	2102	3	1.43	0	0.00	3	1.43	0	0.00	0	0.00	0	0.00	2	0.95
EO_SPEECH_003	1723	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	3	1.74
EO_SPEECH_004	3478	1	0.29	0	0.00	1	0.29	3	0.86	1	0.29	4	1.15	3	0.86
EO_SPEECH_005	2521	1	0.40	0	0.00	1	0.40	0	0.00	0	0.00	0	0.00	12	4.76

EO_SPEECH_006	3122	0	0.00	0	0.00	0	0.00	1	0.32	2	0.64	3	0.96	3	0.96
EO_SPEECH_007	2229	0	0.00	0	0.00	0	0.00	0	0.00	1	0.45	1	0.45	6	2.69
EO_SPEECH_008	2694	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	8	2.97
EO_SPEECH_009	2595	0	0.00	3	1.16	3	1.16	1	0.39	2	0.77	3	1.16	3	1.16
EO_SPEECH_010	2770	2	0.72	1	0.36	3	1.08	1	0.36	3	1.08	4	1.44	5	1.81
EO_SPEECH_011	1848	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	4	2.16
EO_SPEECH_012	2024	1	0.49	0	0.00	1	0.49	0	0.00	2	0.99	2	0.99	3	1.48
EO_SPEECH_013	2659	2	0.75	1	0.38	3	1.13	0	0.00	0	0.00	0	0.00	2	0.75
EO_SPEECH_014	2647	0	0.00	1	0.38	1	0.38	2	0.76	1	0.38	3	1.13	6	2.27
EO_TOU_001	3494	0	0.00	0	0.00	0	0.00	1	0.29	0	0.00	1	0.29	18	5.15
EO_TOU_002	3586	1	0.28	0	0.00	1	0.28	1	0.28	0	0.00	1	0.28	23	6.41
EO_TOU_003	2224	0	0.00	0	0.00	0	0.00	1	0.45	0	0.00	1	0.45	14	6.29
EO_TOU_004	2215	5	2.26	0	0.00	5	2.26	4	1.81	0	0.00	4	1.81	27	12.19
EO_TOU_005	3634	5	1.38	0	0.00	5	1.38	2	0.55	0	0.00	2	0.55	18	4.95
EO_TOU_006	3591	6	1.67	0	0.00	6	1.67	1	0.28	1	0.28	2	0.56	32	8.91
EO_TOU_007	3497	3	0.86	1	0.29	4	1.14	1	0.29	2	0.57	3	0.86	12	3.43
EO_TOU_008	3442	2	0.58	0	0.00	2	0.58	0	0.00	1	0.29	1	0.29	42	12.20
EO_TOU_009	3601	2	0.56	0	0.00	2	0.56	4	1.11	0	0.00	4	1.11	75	20.83
EO_TOU_010	3653	2	0.55	3	0.82	5	1.37	0	0.00	1	0.27	1	0.27	56	15.33
EO_TOU_011	2970	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	19	6.40
EO_WEB_001	3258	0	0.00	0	0.00	0	0.00	2	0.61	1	0.31	3	0.92	18	5.52
EO_WEB_002	3691	0	0.00	0	0.00	0	0.00	0	0.00	2	0.54	2	0.54	21	5.69

EO_WEB_003	4206	0	0.00	2	0.48	2	0.48	1	0.24	2	0.48	3	0.71	55	13.08
EO_WEB_004	1696	1	0.59	0	0.00	1	0.59	0	0.00	2	1.18	2	1.18	5	2.95
EO_WEB_005	2896	1	0.35	2	0.69	3	1.04	0	0.00	1	0.35	1	0.35	45	15.54
EO_WEB_006	1621	1	0.62	1	0.62	2	1.23	0	0.00	0	0.00	0	0.00	28	17.27
EO_WEB_007	3639	1	0.27	0	0.00	1	0.27	1	0.27	1	0.27	2	0.55	27	7.42
EO_WEB_008	2930	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	76	25.94
EO_WEB_009	3618	0	0.00	2	0.55	2	0.55	1	0.28	0	0.00	1	0.28	25	6.91
EO_WEB_010	2546	1	0.39	0	0.00	1	0.39	1	0.39	2	0.79	3	1.18	21	8.25
EO_WEB_011	1933	0	0.00	0	0.00	0	0.00	0	0.00	1	0.52	1	0.52	7	3.62
EO_WEB_012	4085	1	0.24	0	0.00	1	0.24	0	0.00	0	0.00	0	0.00	22	5.39
EO_ACADEMIC_001	3923	11	2.80	0	0.00	11	2.80	0	0.00	1	0.25	1	0.25	4	1.02
EO_ACADEMIC_002	3837	2	0.52	10	2.61	12	3.13	1	0.26	3	0.78	4	1.04	6	1.56
EO_ACADEMIC_003	4317	1	0.23	3	0.69	4	0.93	1	0.23	0	0.00	1	0.23	1	0.23
EO_ACADEMIC_004	3935	0	0.00	0	0.00	0	0.00	0	0.00	1	0.25	1	0.25	0	0.00
EO_ACADEMIC_005	3853	0	0.00	2	0.52	2	0.52	1	0.26	0	0.00	1	0.26	0	0.00
EO_ACADEMIC_006	4093	5	1.22	3	0.73	8	1.95	2	0.49	1	0.24	3	0.73	0	0.00
EO_ACADEMIC_007	3855	2	0.52	0	0.00	2	0.52	1	0.26	0	0.00	1	0.26	0	0.00
EO_ACADEMIC_008	4248	6	1.41	0	0.00	6	1.41	1	0.24	1	0.24	2	0.47	0	0.00
EO_ACADEMIC_009	4239	1	0.24	4	0.94	5	1.18	1	0.24	1	0.24	2	0.47	0	0.00
EO_ACADEMIC_010	4259	1	0.23	6	1.41	7	1.64	2	0.47	1	0.23	3	0.70	0	0.00
EO_FORUM_001	4905	2	0.41	3	0.61	5	1.02	3	0.61	5	1.02	8	1.63	32	6.52
EO_FORUM_002	5353	3	0.56	3	0.56	6	1.12	2	0.37	4	0.75	6	1.12	60	11.21

EO_FORUM_003	3828	0	0.00	4	1.04	4	1.04	0	0.00	2	0.52	2	0.52	23	6.01
EO_FORUM_004	4126	2	0.48	2	0.48	4	0.97	0	0.00	1	0.24	1	0.24	20	4.85
EO_FORUM_005	4102	0	0.00	6	1.46	6	1.46	1	0.24	3	0.73	4	0.98	29	7.07
EO_FORUM_006	3620	6	1.66	1	0.28	7	1.93	1	0.28	1	0.28	2	0.55	20	5.52
EO_FORUM_007	5637	6	1.06	7	1.24	13	2.31	5	0.89	0	0.00	5	0.89	46	8.16
EO_FORUM_008	4073	0	0.00	1	0.25	1	0.25	0	0.00	7	1.72	7	1.72	39	9.58
EO_FORUM_009	3780	2	0.53	1	0.26	3	0.79	4	1.06	0	0.00	4	1.06	50	13.23
EO_FORUM_010	3914	5	1.28	1	0.26	6	1.53	1	0.26	2	0.51	3	0.77	61	15.59
EO_INTERVIEW_001	2091	2	0.96	11	5.26	13	6.22	0	0.00	2	0.96	2	0.96	9	4.30
EO_INTERVIEW_002	3736	1	0.27	6	1.61	7	1.87	0	0.00	2	0.54	2	0.54	8	2.14
EO_INTERVIEW_003	3717	3	0.81	7	1.88	10	2.69	0	0.00	3	0.81	3	0.81	13	3.50
EO_INTERVIEW_004	4558	2	0.44	5	1.10	7	1.54	1	0.22	0	0.00	1	0.22	15	3.29
EO_INTERVIEW_005	3615	3	0.83	12	3.32	15	4.15	0	0.00	5	1.38	5	1.38	24	6.64
EO_INTERVIEW_006	3650	5	1.37	10	2.74	15	4.11	0	0.00	1	0.27	1	0.27	20	5.48
EO_INTERVIEW_007	2252	4	1.78	2	0.89	6	2.66	0	0.00	2	0.89	2	0.89	7	3.11
EO_INTERVIEW_008	3928	2	0.51	2	0.51	4	1.02	1	0.25	2	0.51	3	0.76	1	0.25
EO_INTERVIEW_009	3193	2	0.63	1	0.31	3	0.94	0	0.00	1	0.31	1	0.31	5	1.57
EO_INTERVIEW_010	1974	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	1.01
EO_INTERVIEW_011	3169	6	1.89	11	3.47	17	5.36	0	0.00	4	1.26	4	1.26	4	1.26
EO_INTERVIEW_012	2015	0	0.00	0	0.00	0	0.00	0	0.00	1	0.50	1	0.50	1	0.50
Σ	408016	211		186		397		125		147		272		2168	

Ellipses and fragments in GECCo - English Translations (per text)

TEXT	tokens	nominal coh. a.	nominal coh. n.	verbal / clausal cohesive a.	verbal / clausal cohesive n.	all cohesive a.	all cohesive n.	incoh. nominal a.	incoh. nominal n.	incoh. verbal / clausal a.	incoh. verbal / clausal n.	all incoh. a.	all incoh. n.	other fragments a.	other fragments n.
ETRANS_ESSAY_001	2423	1	0.41	0	0.00	1	0.41	1	0.41	1	0.41	2	0.83	16	6.60
ETRANS_ESSAY_002	1028	0	0.00	0	0.00	0	0.00	1	0.97	0	0.00	1	0.97	9	8.75
ETRANS_ESSAY_003	1881	1	0.53	0	0.00	1	0.53	1	0.53	0	0.00	1	0.53	4	2.13
ETRANS_ESSAY_004	1487	2	1.34	0	0.00	2	1.34	2	1.34	0	0.00	2	1.34	6	4.03
ETRANS_ESSAY_005	1978	0	0.00	0	0.00	0	0.00	1	0.51	0	0.00	1	0.51	5	2.53
ETRANS_ESSAY_006	1911	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	5	2.62
ETRANS_ESSAY_007	2638	0	0.00	2	0.76	2	0.76	1	0.38	0	0.00	1	0.38	13	4.93
ETRANS_ESSAY_008	1110	0	0.00	0	0.00	0	0.00	1	0.90	0	0.00	1	0.90	10	9.01
ETRANS_ESSAY_009	2156	2	0.93	0	0.00	2	0.93	0	0.00	0	0.00	0	0.00	7	3.25
ETRANS_ESSAY_010	2852	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	7	2.45
ETRANS_ESSAY_011	2336	1	0.43	0	0.00	1	0.43	2	0.86	0	0.00	2	0.86	3	1.28
ETRANS_ESSAY_012	534	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	1.87
ETRANS_ESSAY_013	2241	0	0.00	0	0.00	0	0.00	1	0.45	0	0.00	1	0.45	9	4.02
ETRANS_ESSAY_014	2088	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	11	5.27
ETRANS_ESSAY_015	1664	2	1.20	0	0.00	2	1.20	0	0.00	1	0.60	1	0.60	12	7.21
ETRANS_ESSAY_016	817	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	5	6.12
ETRANS_ESSAY_017	1772	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	1.13

ETRANS_ESSAY_018	2037	0	0.00	0	0.00	0	0.00
ETRANS_ESSAY_019	752	1	1.33	1	1.33	2	2.66
ETRANS_ESSAY_020	1422	3	2.11	0	0.00	3	2.11
ETRANS_ESSAY_021	2136	0	0.00	0	0.00	0	0.00
ETRANS_ESSAY_022	2304	0	0.00	0	0.00	0	0.00
ETRANS_ESSAY_023	2469	0	0.00	1	0.41	1	0.41
ETRANS_FICTION_001	4120	3	0.73	8	1.94	11	2.67
ETRANS_FICTION_002	4086	1	0.24	7	1.71	8	1.96
ETRANS_FICTION_003	3971	0	0.00	0	0.00	0	0.00
ETRANS_FICTION_004	4024	1	0.25	0	0.00	1	0.25
ETRANS_FICTION_005	3781	0	0.00	1	0.26	1	0.26
ETRANS_FICTION_006	4548	4	0.88	30	6.60	34	7.48
ETRANS_FICTION_007	3588	1	0.28	4	1.11	5	1.39
ETRANS_FICTION_008	3983	2	0.50	0	0.00	2	0.50
ETRANS_FICTION_009	3948	0	0.00	0	0.00	0	0.00
ETRANS_FICTION_010	3988	7	1.76	10	2.51	17	4.26
ETRANS_INSTR_001	1762	0	0.00	0	0.00	0	0.00
ETRANS_INSTR_002	2266	0	0.00	0	0.00	0	0.00
ETRANS_INSTR_003	4021	0	0.00	0	0.00	0	0.00
ETRANS_INSTR_004	1012	0	0.00	0	0.00	0	0.00
ETRANS_INSTR_005	1615	0	0.00	0	0.00	0	0.00
ETRANS_INSTR_006	2477	0	0.00	0	0.00	0	0.00

0	0.00	0	0.00	0	0.00	4	1.96
0	0.00	1	1.33	1	1.33	5	6.65
1	0.70	0	0.00	1	0.70	6	4.22
1	0.47	1	0.47	2	0.94	14	6.55
0	0.00	1	0.43	1	0.43	6	2.60
0	0.00	1	0.41	1	0.41	8	3.24
3	0.73	3	0.73	6	1.46	37	8.98
0	0.00	5	1.22	5	1.22	25	6.12
2	0.50	3	0.76	5	1.26	4	1.01
3	0.75	1	0.25	4	0.99	13	3.23
0	0.00	1	0.26	1	0.26	33	8.73
8	1.76	0	0.00	8	1.76	40	8.80
0	0.00	0	0.00	0	0.00	39	10.87
1	0.25	2	0.50	3	0.75	60	15.06
0	0.00	1	0.25	1	0.25	7	1.77
2	0.50	2	0.50	4	1.00	50	12.54
0	0.00	0	0.00	0	0.00	20	11.35
0	0.00	0	0.00	0	0.00	40	17.65
0	0.00	0	0.00	0	0.00	40	9.95
0	0.00	0	0.00	0	0.00	30	29.64
0	0.00	0	0.00	0	0.00	33	20.43
0	0.00	0	0.00	0	0.00	35	14.13

ETRANS_INSTR_007	3652	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	31	8.49
ETRANS_INSTR_008	3888	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	58	14.92
ETRANS_INSTR_009	3498	3	0.86	5	1.43	8	2.29	3	0.86	3	0.86	48	13.72
ETRANS_INSTR_010	2551	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	36	14.11
ETRANS_INSTR_011	3542	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	73	20.61
ETRANS_INSTR_012	2097	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	19	9.06
ETRANS_INSTR_013	2925	0	0.00	0	0.00	0	0.00	1	0.34	0	0.00	57	19.49
ETRANS_INSTR_014	4357	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	143	32.82
ETRANS_POPSCI_001	4235	0	0.00	0	0.00	0	0.00	1	0.24	0	0.00	5	1.18
ETRANS_POPSCI_002	4104	1	0.24	2	0.49	3	0.73	1	0.24	0	0.00	10	2.44
ETRANS_POPSCI_003	4085	0	0.00	3	0.73	3	0.73	2	0.49	0	0.00	11	2.69
ETRANS_POPSCI_004	3748	0	0.00	0	0.00	0	0.00	0	0.00	1	0.27	6	1.60
ETRANS_POPSCI_005	3650	0	0.00	1	0.27	1	0.27	1	0.27	0	0.00	7	1.92
ETRANS_POPSCI_006	3640	0	0.00	0	0.00	0	0.00	3	0.82	2	0.55	4	1.10
ETRANS_POPSCI_007	3089	0	0.00	0	0.00	0	0.00	0	0.00	1	0.32	7	2.27
ETRANS_POPSCI_008	4073	0	0.00	1	0.25	1	0.25	1	0.25	0	0.00	12	2.95
ETRANS_POPSCI_009	3952	0	0.00	1	0.25	1	0.25	1	0.25	2	0.51	12	3.04
ETRANS_POPSCI_010	3302	0	0.00	0	0.00	0	0.00	2	0.61	0	0.00	5	1.51
ETRANS_SHARE_001	2250	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	12	5.33
ETRANS_SHARE_002	3027	0	0.00	0	0.00	0	0.00	1	0.33	0	0.00	11	3.63
ETRANS_SHARE_003	4277	0	0.00	0	0.00	0	0.00	1	0.23	0	0.00	9	2.10
ETRANS_SHARE_004	4970	0	0.00	0	0.00	0	0.00	1	0.20	0	0.00	21	4.23

ETRANS_SHARE_005	4657	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	32	6.87
ETRANS_SHARE_006	5400	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	28	5.19
ETRANS_SHARE_007	2703	1	0.37	0	0.00	1	0.37	1	0.37	1	0.37	8	2.96
ETRANS_SHARE_008	3183	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	24	7.54
ETRANS_SHARE_009	3353	0	0.00	0	0.00	0	0.00	1	0.30	1	0.30	8	2.39
ETRANS_SHARE_010	2820	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	15	5.32
ETRANS_SHARE_011	2871	0	0.00	0	0.00	0	0.00	1	0.35	1	0.35	13	4.53
ETRANS_SPEECH_001	1058	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	3	2.84
ETRANS_SPEECH_002	2458	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.41
ETRANS_SPEECH_003	2610	0	0.00	0	0.00	0	0.00	1	0.38	0	0.00	7	2.68
ETRANS_SPEECH_004	2355	2	0.85	0	0.00	2	0.85	1	0.42	0	0.00	9	3.82
ETRANS_SPEECH_005	4014	1	0.25	1	0.25	2	0.50	1	0.25	1	0.25	11	2.74
ETRANS_SPEECH_006	2689	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	3	1.12
ETRANS_SPEECH_007	3159	0	0.00	0	0.00	0	0.00	0	0.00	1	0.32	6	1.90
ETRANS_SPEECH_008	2007	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	5	2.49
ETRANS_SPEECH_009	1467	1	0.68	0	0.00	1	0.68	1	0.68	1	0.68	8	5.45
ETRANS_SPEECH_010	3416	0	0.00	1	0.29	1	0.29	2	0.59	3	0.88	12	3.51
ETRANS_SPEECH_011	803	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	5	6.23
ETRANS_SPEECH_012	3398	2	0.59	0	0.00	2	0.59	0	0.00	1	0.29	3	0.88
ETRANS_SPEECH_013	2714	1	0.37	0	0.00	1	0.37	0	0.00	0	0.00	5	1.84
ETRANS_SPEECH_014	2643	0	0.00	0	0.00	0	0.00	3	1.14	0	0.00	6	2.27
ETRANS_SPEECH_015	1281	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	3	2.34

ETRANS_SPEECH_016	1876	0	0.00	0	0.00	0	0.00
ETRANS_SPEECH_017	805	0	0.00	0	0.00	0	0.00
ETRANS_SPEECH_018	1013	0	0.00	0	0.00	0	0.00
ETRANS_TOU_001	1069	0	0.00	0	0.00	0	0.00
ETRANS_TOU_002	2169	0	0.00	0	0.00	0	0.00
ETRANS_TOU_003	1994	0	0.00	0	0.00	0	0.00
ETRANS_TOU_004	3722	0	0.00	0	0.00	0	0.00
ETRANS_TOU_005	3960	0	0.00	0	0.00	0	0.00
ETRANS_TOU_006	1346	1	0.74	0	0.00	1	0.74
ETRANS_TOU_007	1502	0	0.00	1	0.67	1	0.67
ETRANS_TOU_008	719	0	0.00	0	0.00	0	0.00
ETRANS_TOU_009	3638	1	0.27	0	0.00	1	0.27
ETRANS_TOU_010	2269	0	0.00	0	0.00	0	0.00
ETRANS_TOU_011	2121	3	1.41	0	0.00	3	1.41
ETRANS_TOU_012	2110	0	0.00	0	0.00	0	0.00
ETRANS_TOU_013	780	1	1.28	0	0.00	1	1.28
ETRANS_TOU_014	1111	0	0.00	0	0.00	0	0.00
ETRANS_TOU_015	687	0	0.00	0	0.00	0	0.00
ETRANS_TOU_016	679	2	2.95	0	0.00	2	2.95
ETRANS_TOU_017	628	0	0.00	0	0.00	0	0.00
ETRANS_TOU_018	2043	1	0.49	0	0.00	1	0.49
ETRANS_TOU_019	3391	3	0.88	0	0.00	3	0.88

0	0.00	0	0.00	0	0.00	2	1.07
0	0.00	0	0.00	0	0.00	0	0.00
0	0.00	0	0.00	0	0.00	7	6.91
0	0.00	0	0.00	0	0.00	11	10.29
0	0.00	0	0.00	0	0.00	37	17.06
0	0.00	0	0.00	0	0.00	24	12.04
0	0.00	0	0.00	0	0.00	33	8.87
0	0.00	0	0.00	0	0.00	29	7.32
0	0.00	0	0.00	0	0.00	10	7.43
0	0.00	0	0.00	0	0.00	6	3.99
0	0.00	0	0.00	0	0.00	6	8.34
1	0.27	2	0.55	3	0.82	57	15.67
3	1.32	0	0.00	3	1.32	16	7.05
2	0.94	0	0.00	2	0.94	11	5.19
2	0.95	0	0.00	2	0.95	18	8.53
1	1.28	0	0.00	1	1.28	5	6.41
0	0.00	1	0.90	1	0.90	8	7.20
0	0.00	0	0.00	0	0.00	8	11.64
0	0.00	1	1.47	1	1.47	4	5.89
0	0.00	0	0.00	0	0.00	4	6.37
0	0.00	0	0.00	0	0.00	11	5.38
1	0.29	0	0.00	1	0.29	23	6.78

ETRANS_TOU_020	2172	0	0.00	0	0.00	0	0.00	0	0.00	3	1.38	3	1.38	21	9.67
ETRANS_TOU_021	2361	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	12	5.08
ETRANS_TOU_022	3204	2	0.62	0	0.00	2	0.62	3	0.94	1	0.31	4	1.25	25	7.80
ETRANS_WEB_001	2902	1	0.34	0	0.00	1	0.34	0	0.00	0	0.00	0	0.00	73	25.16
ETRANS_WEB_002	1547	1	0.65	0	0.00	1	0.65	0	0.00	0	0.00	0	0.00	17	10.99
ETRANS_WEB_003	1339	0	0.00	0	0.00	0	0.00	2	1.49	0	0.00	2	1.49	8	5.97
ETRANS_WEB_004	3084	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	20	6.49
ETRANS_WEB_005	4481	1	0.22	0	0.00	1	0.22	2	0.45	0	0.00	2	0.45	53	11.83
ETRANS_WEB_006	4158	2	0.48	0	0.00	2	0.48	0	0.00	1	0.24	1	0.24	51	12.27
ETRANS_WEB_007	1391	0	0.00	0	0.00	0	0.00	0	0.00	1	0.72	1	0.72	57	40.98
ETRANS_WEB_008	4352	1	0.23	0	0.00	1	0.23	1	0.23	0	0.00	1	0.23	20	4.60
ETRANS_WEB_009	4431	1	0.23	0	0.00	1	0.23	0	0.00	0	0.00	0	0.00	24	5.42
ETRANS_WEB_010	2655	2	0.75	0	0.00	2	0.75	0	0.00	0	0.00	0	0.00	26	9.79
ETRANS_WEB_011	3168	0	0.00	1	0.32	1	0.32	0	0.00	0	0.00	0	0.00	24	7.58
ETRANS_WEB_012	3234	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	38	11.75
ETRANS_WEB_013	2915	0	0.00	0	0.00	0	0.00	0	0.00	1	0.34	1	0.34	21	7.20
Σ	322223	67		81		148		75		54		129		2306	

Ellipses and fragments in GECCo - German Originals (per text)

TEXT	tokens	nominal coh. a.	nominal coh. n.	verbal / clausal cohesive a.	verbal / clausal cohesive n.	all cohesive a.	all cohesive n.
GO_ESSAY_001	2035	1	0.49	0	0.00	1	0.49
GO_ESSAY_002	858	1	1.17	0	0.00	1	1.17
GO_ESSAY_003	1554	0	0.00	0	0.00	0	0.00
GO_ESSAY_004	1249	0	0.00	0	0.00	0	0.00
GO_ESSAY_005	1682	1	0.59	0	0.00	1	0.59
GO_ESSAY_006	1653	0	0.00	0	0.00	0	0.00
GO_ESSAY_007	2244	4	1.78	0	0.00	4	1.78
GO_ESSAY_008	971	0	0.00	0	0.00	0	0.00
GO_ESSAY_009	1871	0	0.00	0	0.00	0	0.00
GO_ESSAY_010	2562	0	0.00	0	0.00	0	0.00
GO_ESSAY_011	1787	1	0.56	0	0.00	1	0.56
GO_ESSAY_012	478	0	0.00	0	0.00	0	0.00
GO_ESSAY_013	1986	0	0.00	0	0.00	0	0.00
GO_ESSAY_014	1792	0	0.00	0	0.00	0	0.00
GO_ESSAY_015	1394	2	1.43	0	0.00	2	1.43
GO_ESSAY_016	646	0	0.00	0	0.00	0	0.00
GO_ESSAY_017	1460	0	0.00	0	0.00	0	0.00
GO_ESSAY_018	1782	1	0.56	0	0.00	1	0.56

incoh. nominal a.	incoh. nominal n.	incoh. verbal / clausal a.	incoh. verbal / clausal n.	all incoh. a.	all incoh. n.	other fragments a.	other fragments n.
0	0.00	1	0.49	1	0.49	21	10.32
1	1.17	0	0.00	1	1.17	11	12.82
0	0.00	0	0.00	0	0.00	7	4.50
0	0.00	0	0.00	0	0.00	7	5.60
0	0.00	0	0.00	0	0.00	9	5.35
0	0.00	0	0.00	0	0.00	8	4.84
0	0.00	2	0.89	2	0.89	35	15.60
2	2.06	0	0.00	2	2.06	9	9.27
0	0.00	0	0.00	0	0.00	9	4.81
0	0.00	0	0.00	0	0.00	9	3.51
1	0.56	0	0.00	1	0.56	2	1.12
0	0.00	0	0.00	0	0.00	2	4.18
0	0.00	0	0.00	0	0.00	23	11.58
0	0.00	2	1.12	2	1.12	13	7.25
0	0.00	3	2.15	3	2.15	8	5.74
0	0.00	0	0.00	0	0.00	8	12.38
0	0.00	3	2.05	3	2.05	3	2.05
1	0.56	0	0.00	1	0.56	4	2.24

GO_ESSAY_019	645	1	1.55	2	3.10	3	4.65
GO_ESSAY_020	1220	3	2.46	0	0.00	3	2.46
GO_ESSAY_021	1851	3	1.62	0	0.00	3	1.62
GO_ESSAY_022	1795	0	0.00	0	0.00	0	0.00
GO_ESSAY_023	2153	0	0.00	1	0.46	1	0.46
GO_FICTION_001	3734	1	0.27	7	1.87	8	2.14
GO_FICTION_002	3793	5	1.32	9	2.37	14	3.69
GO_FICTION_003	3675	0	0.00	0	0.00	0	0.00
GO_FICTION_004	3621	0	0.00	0	0.00	0	0.00
GO_FICTION_005	3572	2	0.56	1	0.28	3	0.84
GO_FICTION_006	3989	5	1.25	35	8.77	40	10.03
GO_FICTION_007	3312	1	0.30	0	0.00	1	0.30
GO_FICTION_008	3753	5	1.33	8	2.13	13	3.46
GO_FICTION_009	3582	0	0.00	0	0.00	0	0.00
GO_FICTION_010	3747	7	1.87	10	2.67	17	4.54
GO_INSTR_001	1545	0	0.00	0	0.00	0	0.00
GO_INSTR_002	2282	0	0.00	0	0.00	0	0.00
GO_INSTR_003	3449	0	0.00	0	0.00	0	0.00
GO_INSTR_004	1345	0	0.00	0	0.00	0	0.00
GO_INSTR_005	1433	0	0.00	0	0.00	0	0.00
GO_INSTR_006	2179	0	0.00	0	0.00	0	0.00
GO_INSTR_007	3414	2	0.59	0	0.00	2	0.59

1	1.55	1	1.55	2	3.10	9	13.95
2	1.64	0	0.00	2	1.64	6	4.92
1	0.54	2	1.08	3	1.62	11	5.94
0	0.00	1	0.56	1	0.56	3	1.67
0	0.00	0	0.00	0	0.00	10	4.64
3	0.80	1	0.27	4	1.07	39	10.44
1	0.26	5	1.32	6	1.58	30	7.91
2	0.54	2	0.54	4	1.09	1	0.27
4	1.10	1	0.28	5	1.38	6	1.66
1	0.28	2	0.56	3	0.84	37	10.36
4	1.00	0	0.00	4	1.00	62	15.54
0	0.00	0	0.00	0	0.00	53	16.00
4	1.07	2	0.53	6	1.60	68	18.12
0	0.00	1	0.28	1	0.28	10	2.79
4	1.07	2	0.53	6	1.60	55	14.68
0	0.00	0	0.00	0	0.00	24	15.53
0	0.00	0	0.00	0	0.00	39	17.09
0	0.00	0	0.00	0	0.00	39	11.31
0	0.00	0	0.00	0	0.00	89	66.17
0	0.00	0	0.00	0	0.00	37	25.82
2	0.92	0	0.00	2	0.92	37	16.98
1	0.29	1	0.29	2	0.59	39	11.42

GO_INSTR_008	3487	1	0.29	0	0.00	1	0.29
GO_INSTR_009	3529	4	1.13	4	1.13	8	2.27
GO_INSTR_010	2563	0	0.00	0	0.00	0	0.00
GO_INSTR_011	3466	0	0.00	0	0.00	0	0.00
GO_INSTR_012	1954	0	0.00	0	0.00	0	0.00
GO_INSTR_013	2597	0	0.00	0	0.00	0	0.00
GO_INSTR_014	3637	0	0.00	0	0.00	0	0.00
GO_POPSCI_001	3569	1	0.28	0	0.00	1	0.28
GO_POPSCI_002	3729	0	0.00	2	0.54	2	0.54
GO_POPSCI_003	3731	3	0.80	3	0.80	6	1.61
GO_POPSCI_004	3567	0	0.00	0	0.00	0	0.00
GO_POPSCI_005	3666	3	0.82	0	0.00	3	0.82
GO_POPSCI_006	3619	2	0.55	0	0.00	2	0.55
GO_POPSCI_007	3595	0	0.00	1	0.28	1	0.28
GO_POPSCI_008	3628	1	0.28	1	0.28	2	0.55
GO_POPSCI_009	3538	0	0.00	1	0.28	1	0.28
GO_POPSCI_010	3535	0	0.00	0	0.00	0	0.00
GO_SHARE_001	1914	0	0.00	0	0.00	0	0.00
GO_SHARE_002	2554	0	0.00	0	0.00	0	0.00
GO_SHARE_003	4059	2	0.49	0	0.00	2	0.49
GO_SHARE_004	4288	0	0.00	0	0.00	0	0.00
GO_SHARE_005	4157	0	0.00	0	0.00	0	0.00

2	0.57	0	0.00	2	0.57	143	41.01
0	0.00	3	0.85	3	0.85	85	24.09
0	0.00	1	0.39	1	0.39	34	13.27
1	0.29	0	0.00	1	0.29	87	25.10
0	0.00	0	0.00	0	0.00	22	11.26
0	0.00	0	0.00	0	0.00	120	46.21
0	0.00	0	0.00	0	0.00	167	45.92
1	0.28	0	0.00	1	0.28	4	1.12
3	0.80	0	0.00	3	0.80	22	5.90
3	0.80	2	0.54	5	1.34	16	4.29
1	0.28	1	0.28	2	0.56	6	1.68
1	0.27	0	0.00	1	0.27	9	2.45
0	0.00	2	0.55	2	0.55	5	1.38
3	0.83	1	0.28	4	1.11	3	0.83
5	1.38	0	0.00	5	1.38	10	2.76
1	0.28	1	0.28	2	0.57	12	3.39
1	0.28	0	0.00	1	0.28	4	1.13
0	0.00	0	0.00	0	0.00	8	4.18
0	0.00	0	0.00	0	0.00	7	2.74
0	0.00	0	0.00	0	0.00	9	2.22
0	0.00	0	0.00	0	0.00	19	4.43
0	0.00	0	0.00	0	0.00	22	5.29

GO_SHARE_006	4988	0	0.00	0	0.00	0	0.00
GO_SHARE_007	2620	0	0.00	0	0.00	0	0.00
GO_SHARE_008	2762	0	0.00	0	0.00	0	0.00
GO_SHARE_009	3105	0	0.00	0	0.00	0	0.00
GO_SHARE_010	2421	0	0.00	0	0.00	0	0.00
GO_SHARE_011	2367	0	0.00	0	0.00	0	0.00
GO_SPEECH_001	874	0	0.00	0	0.00	0	0.00
GO_SPEECH_002	1982	0	0.00	0	0.00	0	0.00
GO_SPEECH_003	2355	0	0.00	0	0.00	0	0.00
GO_SPEECH_004	1890	2	1.06	0	0.00	2	1.06
GO_SPEECH_005	3158	2	0.63	1	0.32	3	0.95
GO_SPEECH_006	2747	0	0.00	0	0.00	0	0.00
GO_SPEECH_007	2817	0	0.00	1	0.35	1	0.35
GO_SPEECH_008	1882	1	0.53	0	0.00	1	0.53
GO_SPEECH_009	1267	2	1.58	0	0.00	2	1.58
GO_SPEECH_010	2973	0	0.00	1	0.34	1	0.34
GO_SPEECH_011	718	0	0.00	0	0.00	0	0.00
GO_SPEECH_012	3117	2	0.64	0	0.00	2	0.64
GO_SPEECH_013	2809	2	0.71	0	0.00	2	0.71
GO_SPEECH_014	2535	0	0.00	0	0.00	0	0.00
GO_SPEECH_015	1046	0	0.00	0	0.00	0	0.00
GO_SPEECH_016	1655	0	0.00	0	0.00	0	0.00

1	0.20	0	0.00	1	0.20	25	5.01
0	0.00	0	0.00	0	0.00	10	3.82
0	0.00	0	0.00	0	0.00	8	2.90
1	0.32	0	0.00	1	0.32	8	2.58
1	0.41	0	0.00	1	0.41	18	7.43
0	0.00	1	0.42	1	0.42	12	5.07
0	0.00	0	0.00	0	0.00	6	6.86
2	1.01	0	0.00	2	1.01	6	3.03
0	0.00	0	0.00	0	0.00	1	0.42
1	0.53	0	0.00	1	0.53	9	4.76
1	0.32	1	0.32	2	0.63	25	7.92
0	0.00	0	0.00	0	0.00	10	3.64
1	0.35	1	0.35	2	0.71	12	4.26
0	0.00	0	0.00	0	0.00	7	3.72
0	0.00	0	0.00	0	0.00	10	7.89
1	0.34	2	0.67	3	1.01	17	5.72
0	0.00	0	0.00	0	0.00	3	4.18
1	0.32	0	0.00	1	0.32	9	2.89
2	0.71	2	0.71	4	1.42	12	4.27
2	0.79	0	0.00	2	0.79	20	7.89
1	0.96	0	0.00	1	0.96	6	5.74
0	0.00	0	0.00	0	0.00	2	1.21

GO_SPEECH_017	640	0	0.00	0	0.00	0	0.00
GO_SPEECH_018	934	0	0.00	0	0.00	0	0.00
GO_TOU_001	818	0	0.00	0	0.00	0	0.00
GO_TOU_002	1722	0	0.00	0	0.00	0	0.00
GO_TOU_003	1631	0	0.00	0	0.00	0	0.00
GO_TOU_004	3457	0	0.00	0	0.00	0	0.00
GO_TOU_005	3506	0	0.00	0	0.00	0	0.00
GO_TOU_006	1151	0	0.00	0	0.00	0	0.00
GO_TOU_007	1258	0	0.00	1	0.79	1	0.79
GO_TOU_008	577	1	1.73	0	0.00	1	1.73
GO_TOU_009	2662	0	0.00	0	0.00	0	0.00
GO_TOU_010	1883	0	0.00	0	0.00	0	0.00
GO_TOU_011	1754	3	1.71	0	0.00	3	1.71
GO_TOU_012	1823	1	0.55	0	0.00	1	0.55
GO_TOU_013	658	1	1.52	0	0.00	1	1.52
GO_TOU_014	879	1	1.14	0	0.00	1	1.14
GO_TOU_015	561	2	3.57	0	0.00	2	3.57
GO_TOU_016	552	1	1.81	0	0.00	1	1.81
GO_TOU_017	516	1	1.94	0	0.00	1	1.94
GO_TOU_018	1806	1	0.55	0	0.00	1	0.55
GO_TOU_019	2831	3	1.06	0	0.00	3	1.06
GO_TOU_020	1779	0	0.00	0	0.00	0	0.00

0	0.00	1	1.56	1	1.56	0	0.00
0	0.00	0	0.00	0	0.00	6	6.42
0	0.00	0	0.00	0	0.00	3	3.67
0	0.00	0	0.00	0	0.00	47	27.29
0	0.00	0	0.00	0	0.00	16	9.81
0	0.00	4	1.16	4	1.16	29	8.39
0	0.00	1	0.29	1	0.29	28	7.99
0	0.00	0	0.00	0	0.00	8	6.95
1	0.79	1	0.79	2	1.59	12	9.54
0	0.00	0	0.00	0	0.00	7	12.13
1	0.38	1	0.38	2	0.75	54	20.29
1	0.53	0	0.00	1	0.53	20	10.62
1	0.57	1	0.57	2	1.14	13	7.41
0	0.00	1	0.55	1	0.55	25	13.71
0	0.00	1	1.52	1	1.52	7	10.64
0	0.00	1	1.14	1	1.14	12	13.65
0	0.00	0	0.00	0	0.00	9	16.04
0	0.00	2	3.62	2	3.62	3	5.43
0	0.00	0	0.00	0	0.00	4	7.75
0	0.00	0	0.00	0	0.00	11	6.09
0	0.00	3	1.06	3	1.06	30	10.60
0	0.00	3	1.69	3	1.69	31	17.43

GO_TOU_021	2314	1	0.43	0	0.00	1	0.43
GO_TOU_022	2436	1	0.41	0	0.00	1	0.41
GO_WEB_001	2640	1	0.38	0	0.00	1	0.38
GO_WEB_002	1252	0	0.00	0	0.00	0	0.00
GO_WEB_003	1277	0	0.00	0	0.00	0	0.00
GO_WEB_004	2834	3	1.06	0	0.00	3	1.06
GO_WEB_005	4079	0	0.00	0	0.00	0	0.00
GO_WEB_006	3401	0	0.00	0	0.00	0	0.00
GO_WEB_007	1185	0	0.00	0	0.00	0	0.00
GO_WEB_008	3959	1	0.25	0	0.00	1	0.25
GO_WEB_009	3652	0	0.00	0	0.00	0	0.00
GO_WEB_010	2918	1	0.34	0	0.00	1	0.34
GO_WEB_011	3354	0	0.00	1	0.30	1	0.30
GO_WEB_012	2656	0	0.00	0	0.00	0	0.00
GO_WEB_013	2572	0	0.00	0	0.00	0	0.00
GO_ACADEMIC_001	3784	1	0.26	3	0.79	4	1.06
GO_ACADEMIC_002	3785	2	0.53	2	0.53	4	1.06
GO_ACADEMIC_003	3547	6	1.69	3	0.85	9	2.54
GO_ACADEMIC_004	10131	2	0.20	2	0.20	4	0.39
GO_ACADEMIC_005	3271	3	0.92	1	0.31	4	1.22
GO_ACADEMIC_006	3964	13	3.28	10	2.52	23	5.80
GO_ACADEMIC_007	4066	7	1.72	7	1.72	14	3.44

0	0.00	0	0.00	0	0.00	12	5.19
0	0.00	2	0.82	2	0.82	23	9.44
0	0.00	1	0.38	1	0.38	47	17.80
1	0.80	0	0.00	1	0.80	0	0.00
2	1.57	0	0.00	2	1.57	12	9.40
0	0.00	0	0.00	0	0.00	20	7.06
1	0.25	3	0.74	4	0.98	48	11.77
1	0.29	0	0.00	1	0.29	33	9.70
0	0.00	0	0.00	0	0.00	42	35.44
1	0.25	0	0.00	1	0.25	23	5.81
2	0.55	0	0.00	2	0.55	18	4.93
0	0.00	1	0.34	1	0.34	28	9.60
0	0.00	1	0.30	1	0.30	20	5.96
1	0.38	0	0.00	1	0.38	31	11.67
0	0.00	0	0.00	0	0.00	20	7.78
2	0.53	1	0.26	3	0.79	3	0.79
0	0.00	0	0.00	0	0.00	1	0.26
3	0.85	3	0.85	6	1.69	14	3.95
2	0.20	1	0.10	3	0.30	50	4.94
1	0.31	2	0.61	3	0.92	1	0.31
4	1.01	0	0.00	4	1.01	4	1.01
2	0.49	4	0.98	6	1.48	14	3.44

GO_ACADEMIC_008	3815	4	1.05	0	0.00	4	1.05
GO_ACADEMIC_009	3790	0	0.00	0	0.00	0	0.00
GO_ACADEMIC_010	3550	2	0.56	0	0.00	2	0.56
GO_FORUM_001	4532	3	0.66	5	1.10	8	1.77
GO_FORUM_002	3879	4	1.03	2	0.52	6	1.55
GO_FORUM_003	5415	2	0.37	6	1.11	8	1.48
GO_FORUM_004	3876	1	0.26	6	1.55	7	1.81
GO_FORUM_005	3570	0	0.00	0	0.00	0	0.00
GO_FORUM_006	4269	17	3.98	0	0.00	17	3.98
GO_FORUM_007	3798	1	0.26	5	1.32	6	1.58
GO_FORUM_008	4267	0	0.00	0	0.00	0	0.00
GO_FORUM_009	3910	6	1.53	15	3.84	21	5.37
GO_FORUM_010	4120	5	1.21	4	0.97	9	2.18
GO_INTERVIEW_001	3313	0	0.00	4	1.21	4	1.21
GO_INTERVIEW_002	4028	4	0.99	1	0.25	5	1.24
GO_INTERVIEW_003	4965	1	0.20	7	1.41	8	1.61
GO_INTERVIEW_004	1789	2	1.12	1	0.56	3	1.68
GO_INTERVIEW_005	3307	4	1.21	2	0.60	6	1.81
GO_INTERVIEW_006	3112	0	0.00	1	0.32	1	0.32
GO_INTERVIEW_007	2754	0	0.00	1	0.36	1	0.36
GO_INTERVIEW_008	3745	6	1.60	0	0.00	6	1.60
GO_INTERVIEW_009	2994	4	1.34	4	1.34	8	2.67

2	0.52	0	0.00	2	0.52	6	1.57
1	0.26	0	0.00	1	0.26	1	0.26
0	0.00	0	0.00	0	0.00	0	0.00
2	0.44	6	1.32	8	1.77	13	2.87
1	0.26	2	0.52	3	0.77	40	10.31
1	0.18	1	0.18	2	0.37	27	4.99
2	0.52	2	0.52	4	1.03	43	11.09
2	0.56	1	0.28	3	0.84	23	6.44
2	0.47	4	0.94	6	1.41	25	5.86
5	1.32	0	0.00	5	1.32	30	7.90
3	0.70	1	0.23	4	0.94	21	4.92
3	0.77	2	0.51	5	1.28	34	8.70
1	0.24	3	0.73	4	0.97	27	6.55
0	0.00	0	0.00	0	0.00	4	1.21
0	0.00	0	0.00	0	0.00	7	1.74
0	0.00	2	0.40	2	0.40	8	1.61
0	0.00	0	0.00	0	0.00	7	3.91
0	0.00	0	0.00	0	0.00	30	9.07
0	0.00	0	0.00	0	0.00	18	5.78
1	0.36	0	0.00	1	0.36	4	1.45
0	0.00	0	0.00	0	0.00	3	0.80
0	0.00	0	0.00	0	0.00	2	0.67

GO_INTERVIEW_010	2687	0	0.00	2	0.74	2	0.74
GO_INTERVIEW_011	1231	0	0.00	0	0.00	0	0.00
GO_INTERVIEW_012	1803	2	1.11	1	0.55	3	1.66
GO_INTERVIEW_013	1417	6	4.23	5	3.53	11	7.76
GO_INTERVIEW_014	3053	2	0.66	0	0.00	2	0.66
Σ	414027	207		190		397	

0	0.00	2	0.74	2	0.74	1	0.37
0	0.00	0	0.00	0	0.00	1	0.81
1	0.55	1	0.55	2	1.11	8	4.44
0	0.00	0	0.00	0	0.00	9	6.35
0	0.00	0	0.00	0	0.00	8	2.62
126		116		242		3181	