Simplification in simultaneous interpreting - a lexical and syntactic perspective

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

vorgelegt von
Heike Przybyl
aus Tübingen

Saarbrücken, 26. Februar 2026

Dekanin der Fakultät P: Prof. Dr. Nine Miedema
Erstberichterstatterin: Prof. Dr. Elke Teich
Zweitberichterstatterin: Prof. Dr. Ekaterina Lapshinova-Koltunski
Tag der letzten Prüfungsleistung: 26. Juni 2026

Acknowledgements

I would like to express my heartfelt gratitude to the several individuals and institutions who supported and patiently endured me during the long process of completing my dissertation.

First and foremost, I would like to express my sincere gratitude to my supervisors, Prof. Elke Teich and Prof. Ekaterina Lapshinova-Koltunski, for their invaluable support and guidance, and especially for their patience. Their expertise, thoughtful feedback, and encouragement have been instrumental throughout this project.

I would also like to extend my sincere thanks to the B7 project team, especially Maria and Christina, for their in-depth discussions on the subject matter and their valuable insights. It has been a pleasure and a privilege to work alongside them. Thanks also to the student assistants who supported the corpus compilation process.

Furthermore, I would also like to thank my colleagues in our department, especially Stefan and Jörg, for their invaluable help in solving the many technical challenges that arose throughout this project. Their expertise and support were indispensable during the corpus compilation and annotation process. I am equally grateful to Pauline, Koel, Stefania, Katrin, Mihaela, Marie-Ann, Sabine and all the colleagues on our corridor for their encouragement and professional support, all of which made this journey both more enjoyable and more rewarding.

I would like to acknowledge the support provided by SFB 1102 in supporting my research and in providing the necessary infrastructure for carrying out this study. Thanks to Saarland University and the Language Science and Technology Department for providing a supportive academic environment and access to the necessary resources for my research.

Finally, I would like to express my deepest gratitude to my family for their unwavering support throughout this long journey.

Completing this dissertation has been a challenging yet rewarding endeavour, and I could not have done it without your support. I am deeply grateful for the trust you placed in me and for your contributions to this work. Thank you all for being part of this academic journey.

Abstract

Simultaneous interpreting is a language production mode that is influenced by the particular constraints of the interpreting process. This thesis provides a profound understanding of the specific linguistic properties of interpreted language, also referred to as *interpretese*. It investigates the idea that, due to the overall cognitively challenging task of simultaneous interpreting, which includes simultaneously coordinating decoding of source input, encoding and monitoring of target output, intermediately storing source and target items in memory until production is completed (Gile, 1995), specific traces of optimal encoding will be visible in the interpreting product. The main perspective taken in this thesis is rational communication, assuming that speakers strive to transmit the source message accurately while optimising their cognitive effort. In simultaneous interpreting, challenging production conditions influence rational choices that are expected to result in simplification of the interpreting product. This hypothesis is studied through the lexical and syntactic choices made by the interpreters. To investigate the simplification hypothesis, this thesis takes a corpus-based approach and investigates a high quality interpreting corpus, EPIC-UdS (Przybyl et al., 2022b). It uses a comparable and parallel approach to the German and English datasets to investigate (a) simplification in interpreting compared to target language norms and (b) simplification interpreting compared to source language input.

As a unique approach combining predictability and memory, this thesis uses measures that have so far not been applied in interpreting research, acknowledging that language processing in general, and also in challenging production conditions such as interpreting, involves optimal encoding and memory optimisation. First, the thesis uses the information-theoretic measure of relative entropy (Kullback-Leibler divergence, KLD) (Kullback and Leibler, 1951) to detect distinctive features of simultaneous interpreting compared to original spoken language and gives a data-driven description of simultaneous interpreted language without prior selection of features to be investigated. To account for predictability and memory, this thesis uses measures to investigate simplification that provide a direct link to cognition. This particularly includes the second framework used in this thesis. The information-theoretic measure of surprisal, or information content encoded in a linguistic unit (Hale, 2001), is used to account for predictability at the lexical level. Surprisal will be high when a word in its context has a low probability. It can be seen as an indicator of the effort required to comprehend or produce language (Levy, 2008; Smith and Levy, 2013; Bicknell and Levy, 2010; Demberg and Keller, 2008). More expected words, i.e. words with lower surprisal are linked to a higher degree of simplification and expected language use, while less expected words are linked to less simplification. This corpus-based approach to cognitive processing that uses predictability of words in context is complemented by corpus-based measures to investigate lexical simplification that were traditionally used in simplification research in interpreting studies (Laviosa, 1998; Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2012, 2015; Ferraresi et al., 2018; Dayter, 2018).

As third framework used in this thesis, memory optimisation is approach via Dependency Locality Theory (DLT). Syntactic simplification is investigated as the linear distance between syntactically related words, which can be linked to the effort to maintain a linguistic item in working memory (Gibson, 2000; Liu, 2008). Shorter distances between syntactic elements are linked to lower syntactic complexity. Minimisation of dependency length (DLM) is generally observed in natural languages (Gildea and Temperley, 2010; Temperley, 2007) and is investigated here for simultaneous interpreting due to the assumed pressure to optimise memory use. This syntactic approach to simplification is combined with a source-to-target alignment analysis at the sentence level in order to investigate whether and if so, how interpreters deal with source speech input at the sentence level.

The results can be summarised as follows. *Interpretese* is particularly characterised by elements of spoken language and prefers to use verbal rather than nominal elements. This is in line with previous studies that include other language combinations (Shlesinger and Ordan, 2012; Karakanta et al., 2021; Gast and Borges, 2023). Previous findings on lexical simplification for other language combinations using traditional measures were mostly confirmed (Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2015; Ferraresi et al., 2018; Dayter, 2018; Kajzer-Wietrzny and Ivaska, 2020), while this thesis provides more fine-grained measures using annotation on delivery mode and rate. Moreover, lexical simplification is also generally confirmed via the use of more expected words in context in interpreting, i.e. the use of lower surprisal words. At the syntactic level, no general trend of (DLM) is observed, but interpreters rather simplify by producing shorter sentences, for example, by splitting source sentences into several target sentences. Besides simplification, cognitive facilitation is achieved by staying close to the source input, syntactically as well as lexically, through the use of cognates, i.e. words in the target that are lexically very close to the source speech equivalent. Generally, the main assumption that the cognitively challenging production condition in interpreting would require interpreters to encode their output optimally and that this would leave traces of lexical and syntactic simplification in the interpreting product is largely confirmed, although more so in English than in German.

Predictability and memory optimisation lead to optimised output in monolingual production conditions. This thesis showed that in the particularly contained production setting of simultaneous interpreting, interpreters efficiently choose and order words. Optimal encoding in this setting can be linked to simplifications trends, but is also influenced by source language shining-through as a trace of memory optimisation in simultaneous interpreting.

Contents

List of Abbreviations

ADJ adjective

ADP adposition

ADV adverb

AUX auxiliary

av DL average dependency length

av max DL average maximum dependency length

AvS average surprisal

CBIS corpus-based interpreting studies

CCONJ coordinating conjunction

CLM Cognitive Load Model

DD dependency distance

DET determiner

DL dependency length

DLM Dependency Length Minimisation

DLT Dependency Locality Theory

EP European Parliament

EPIC-UdS European Parliament Interpreting Corpus at Saarland University

EVS ear-voice span

GAMM generalized additive mixed-effects model

INTJ interjection

KLD Kullback-Leibler divergence

LD lexical density

LM language model

max DD maximum dependency distance

MEPs members of the European Parliament

MU memory unit

NMT neural machine translation

NOUN noun

ORG DE original spoken German

ORG EN original spoken English

POS part-of-speech

PRON pronoun

PROPN proper noun

RQs research questions

SCONJ subordinating conjunction

SI DE EN simultaneous interpreting from German into English

SI EN DE simultaneous interpreting from English into German

SI ES EN simultaneous interpreting from Spanish into English

SL source language

ST source text

STTR standardised type-token ratio

TEPRs task-evoked pupillary responses

TL target language

TT target text

TTR type-token ratio

VERB verb

wpm words per minute

List of Figures

List of Tables

Part I

Introduction and background

Chapter 1

Introduction

This thesis studies the specific properties of simultaneous interpreting. It links *interpretese* to cognition using a corpus-based approach to study *simplification* with information-theoretic measures.

Research in corpus-based translation studies has shown that written translations differ significantly from comparable original production. These differences can be grouped into lexicogrammatical, morphosyntactic, and textual language patterns and are referred to as *translationese* (Gellerstam, 1986). There is a tendency of translations to be more explicit, more implicit, or more simplified than original texts/source texts (Toury, 1995; Olohan and Baker, 2000). The translation process can also lead to target texts that overuse target language conventions (normalisation) or still show traces of the source texts/languages they are based on (shining-through) (Teich, 2003). The universality of these translationese categories has been disputed (House, 2016), however, the existence of differences between originals and translated language cannot be denied. Machines can reliably distinguish translations from comparable original production (Baroni and Bernardini, 2006; Volansky et al., 2015; Rubino et al., 2016; Kunilovskaya and Lapshinova-Koltunski, 2020).

Translationese for written translations is widely explored in the literature, however, few studies deal with the specific properties of other forms of mediated communication such as interpreting. Simultaneous interpreting is a spoken mode of translation, but the interpreting process differs greatly from the written translation process. Although time constraints certainly also exist for written translations, the time aspect plays an essential role in simultaneous interpreting. Incoming source speech has to be decoded, intermediately stored, and encoded by the interpreter almost instantly. Additionally, the interpreter needs to monitor their own output and make corrections if necessary. Usually, this happens within 1 to 3 seconds and can be measured with a span of a source element uttered and the corresponding target pronounced (Defrancq, 2015). This almost instant translation process with no possibility of pausing the process or relistening to a source speech segment can be expected to leave particular traces in the

interpreting product, which may differ from observed translationese characteristics.

Simultaneous interpreting research acknowledges the particular role of cognitive constraints in simultaneous interpreting. Many theoretical interpreting models try to capture the interpreting process and show how limited cognitive resources are managed (Gerver, 1976; Moser, 1978; Gile, 1995, 2009, 2017; Seeber, 2011, 2013). While they emphasise different aspects of the interpreting process, they all see particular implications for interpreter's (short-term) memory and high cognitive demand. Unfortunately, these models are only of theoretical nature, and apart from Seeber's cognitive load model (Seeber, 2011; Seeber and Kerzel, 2012b), these models are not designed to be tested empirically. Therefore, they cannot be validated in a real interpreting context. More recently, experimental approaches have attempted to capture the cognitive load during the interpreting process via pupillometry (Seeber, 2013) or retrospective protocols (Gumul, 2017, 2020; Ivanova, 2000; Bartłomiejczyk, 2006). Both approaches, as well as the approach to use corpus data and look at disfluencies as signs of high cognitive load. They use potential triggers of disfluencies, such as high lexical density, delivery rate, and numbers, as predictor variables for disfluencies in the interpreting product and focus on peaks of cognitive load.

This thesis also uses a corpus-based approach to investigate the effects of cognitive constraints. However, traditional corpus-based approaches to capture specific characteristics of simultaneous interpreting, investigating *interpretese*, only provide a descriptive perspective and give a mixed picture (Kajzer-Wietrzny, 2015; Sandrelli and Bendazzoli, 2005; Ferraresi et al., 2018; Dayter, 2018; Lv and Liang, 2019). The perspective taken in this thesis differs from previous corpus-based approaches that focus on problem triggers which may result in high cognitive load. This thesis rather assumes that due to the complexity of the task overall, interpreters, just as original speakers, are required to encode their output optimally. In order to cope with restricted time and cognitive resources, the interpreter may use the linguistic options available to them to optimally encode the message overall, which is expected to be simpler encoding than in non-constraint communication. Natural languages have shown a tendency to encode messages efficiently (Gibson et al., 2019). This thesis investigates whether the effects of the challenging productions conditions can be found in the interpreting product as an optimal encoding of the message. Optimal encoding in interpreting may be a production-oriented mechanism, whereas original speakers who aim to encode their message optimally are found to do so as audience design.

What we know about properties of *interpretese* is still fairly limited. Previous studies investigating interpreted language have shown that the translationese characteristics found to be valid for written translation do not all apply to spoken translation (Kajzer-Wietrzny, 2012). Simultaneous interpreting shows more characteristics of spoken language than of (written) translation (Shlesinger and Ordan, 2012). Corpus-based interpreting studies (CBIS) only go back some decades. Due to the limited availability of interpreting data and the great time effort required to compile high-quality interpreting corpora, corpus-based interpreting studies are still rare and

restricted to very few language pairs. In addition to not sharing all of the same characteristics of written translations, studies show that traditional corpus-based measures give inconclusive results, depending on the languages and corpus resources studied and measures applied. For example, studies on European Parliament data such as EPIC (Russo et al., 2012) and TIC (Kajzer-Wietrzny, 2012) show higher lexical density and therefore less simplification in interpreted language compared to spoken originals in the same language. An opposite trend was shown for interpreting into Russian (Dayter, 2018), therefore, the languages and language pairs studied make a difference. Generally, reduced syntactic complexity is observed, as indicated by shallow features such as sentence length (Russo et al., 2012) and shorter midfields (Collard et al., 2019). Furthermore, Kajzer-Wietrzny and Grabowski (2021) show reduced lexical variation indicated by more frequent bigrams used in interpreting. In general, there seems to be a preference for a less nominal but rather verbal style in interpreting (Gast and Borges, 2023; Shlesinger and Ordan, 2012). Nouns used in interpreting are generally drawn from the same vocabulary as in comparable original speeches; however, interpreters tend to use generic nouns more frequently (Bizzoni et al., 2021).

Generally, a general tendency towards reduced variation in interpreting seems to be observed, using traditional corpus-based and statistical methods. Using language model (LM) trained on a corpus with readability annotations and subsequently performing a number of classification and regression tasks, Kunilovskaya et al. (2023a) find simultaneous interpreting to be more readable and therefore more simple than comparable originals, as well as comparable written translations in the same language. They use the same corpus that is used in this thesis, EPIC-UdS, but apply a different set of measures and particularly focus on complex and simple sentences.

What we know empirically about *interpretese* is very limited, and using traditional corpus-based methods that may work for written translations does not capture the cognitive aspects of language production. The characteristics that define interpreted language may potentially intersect with the trends found for translation. Considering the particular time and cognitive constraints of the interpreting process, as well as spoken language phenomena, interpreters may not have the additional free cognitive resources to focus on audience design and try to make things more explicit (as translators are found to do). They may rather aim to manage available resources efficiently at all times by encoding their output in an optimal way and simplify by reducing lexical variation, using expected lexis, or by reducing syntactic complexity.

This thesis employs new methods to study *interpretese* and *simplification* in particular: a framework that captures efficient encoding and communication of information is Information Theory as introduced by Shannon (1948). It describes efficient transmission of information across an imperfect noisy channel by quantifying the information content of a message in bits, and allows for comparing different data subsets (in the case of this thesis different subcorpora of EPIC-UdS). The method of *relative entropy (Kullback–Leibler divergence, KLD)* can be used to detect the distinctive

features of a linguistic subset vs another. In this thesis, it is applied to describe the distinctive characteristics of simultaneous interpreting when compared to original spoken language (cf. Chapter 4). Its contribution is a data-driven description of simultaneously interpreted language and therefore adds knowledge towards a definition of *interpretese*. In addition to using traditional corpus-based methods, this thesis employs a further information-theoretic measure. *Surprisal* is considered to be an indicator of the effort required to comprehend or produce language, providing a direct link to cognition (Hale, 2001; Teich et al., 2020). Surprisal measures directly correlate with reading times (Smith and Levy, 2013). While there are some studies in the translation background that also employ information-theoretic measures (Bizzoni and Lapshinova-Koltunski, 2021; Lapshinova-Koltunski et al., 2022; Martínez and Teich, 2017; Rubino et al., 2016; Teich et al., 2020), this work focusses entirely on the spoken mode: simultaneous interpreting and comparable as well all parallel source texts. It investigates the information content (surprisal/expectedness) of distinctive parts-of-speech for interpreting, mainly content words as the main carrier of a message. More expected words, i.e. lower surprisal, are seen as an indicator of more simplification.

A second framework that provides a link to linguistic processing and memory is *Dependency Locality Theory (DLT)* (Gibson, 2000). The linear distance between syntactically related words can be seen as an effort to maintain a linguistic item in working memory. Dependency locality measures can be taken as a proxy of the sentence processing difficulty (Gibson, 2000; Liu, 2008). As with surprisal measures, dependency locality measures also correlate with reading times (Grodner and Gibson, 2005), providing a direct link to cognition. Long distances between syntactically related words may lead to comprehension difficulties (Grodner and Gibson, 2005; Bartek et al., 2011). The tendency to minimise dependency distances as an efficient way of communication can be observed across many languages (Futrell et al., 2015). Interpreters trying to manage their cognitive resources in an optimal way may be expected to minimise dependency distances if possible. However, as they rely on source speech input and cannot influence the message they are supposed to translate, it has to be explored whether the source speech effect in the target (shining-though) will counterbalance such a possible minimisation effect. This thesis uses the Dependency Locality Theory framework to approximate simplification, defining shorter dependency distances and reduced syntactic complexity as more simplified.

As laid out above, due to the particular time and cognitive constraints of the interpreting process, this thesis explores whether features of simplification can be found in the interpreting product. The basic definition of simplification is taken from translation studies, where it is defined as "the process and/or result of making do with less words" (Blum and Levenston, 1978, p.399). "Less words" in this respect may, on one hand, mean shorter encoding, i.e. shorter sentences, less complex syntactic structures, or shorter dependency distances; on the other hand, it may refer to fewer different words and therefore less variation in linguistic choice. This will be measured

with traditional corpus-based methods, such as lexical density and type-token ratio. However, the main focus studying lexical variation will be on surprisal. Expected words in context are low in surprisal due to low variation in the context in which they occur. The focus of studying syntactic complexity will be on dependency measures.

If simplification of the interpreting product can be empirically validated overall, this does not generally reject the existence of source language traces in the interpreting product. Shining-through may still be traceable, potentially manifested in the interpreting product as peaks in cognitive load or the result of high load over a certain stretch of the speech. Lost content or unfinished sentences in the interpreting product are obvious signals of a problem during the interpreting process (which may originate in cognitive overload). They may lead to very unexpected output. Furthermore, traces of the original speech shining-through (cf. translation universals) can also be expected, which may also lead to peaks in surprisal and potentially long dependency distances. In the simultaneous interpreting setting, the interpreter generally cannot influence delivery mode (preparedness of speech), delivery rate (words pronounced per minute), or other characteristics of the source speech. This thesis assumes that the interpreter may use the different linguistic options available to them to encode their output in an optimal way in order to deal with the specific (time) constraints of the interpreting process.

To investigate *interpretese* and the simplification hypothesis in particular, this thesis uses the language pair English and German of the EPIC-UdS corpus (Przybyl et al., 2022b), a European Parliament simultaneous interpreting corpus. It takes a comparable approach, comparing originals to interpreted speeches in the same language (i.e. English or German), as well as a parallel approach, comparing interpreted speeches in the target language to the original speeches they were based on in order to find potential triggers of effect that we see in interpreted speech. Furthermore, the lexical investigations use delivery mode and rate as potential variables influencing the target output.

Overall, the thesis uses information-theoretic measures to investigate (a) how interpreted language differs from original production. It uses *relative entropy* to detect and describe distinctive features of interpretese. In a second step, interpretese features found in this exploratory analysis are used to (b) explore the simplification hypothesis on various levels. The lexical level is analysed via traditional corpus-based methods, investigating lexical variation, as well as the information-theoretic measure of *surprisal*, analysing expectedness in context. The syntactic level is studied via an analysis of the aligned source and target sentences, as well as an investigation of syntactic complexity within the framework of *Dependency Locality Theory*.

As mentioned above, what we empirically know about simultaneous interpreting is still fairly limited. This thesis aims to **link empirical corpus-based investigations into properties of simultaneous interpreting with methods used to approximate cognitive processes novel to corpus-based interpreting studies (CBIS)**. It contributes to the knowledge of *interpretese* by combining traditional

methods with measures that had not been used in interpreting studies until very recently. Therefore, it gives **a more comprehensive overview of the characteristic of interpreted language** than previous work. Furthermore, simultaneous interpreting in German as one of the major spoken languages in the European Parliament has been understudied so far in the literature. By providing the **high-quality corpus resource EPIC-UdS** (Przybyl et al., 2022b), this thesis fills a void by providing a resource that can also be used for studies carried out beyond this thesis. Furthermore, this thesis uses **metadata on delivery rate and mode**, which are also available in some other European Parliament interpreting corpora but not used in a systematic way, and acknowledges the mixed delivery type of this interpreting setting. An analysis of manual **source-to-target sentence alignment** of the entire EPIC-UdS German and English subcorpora provides valuable insight to interpreters' syntactic choices. Beside these contributions, the focus of this thesis is the application of **information-theoretic measures to approximate processing costs** and explain interpreters' linguistic choices in the context of required optimal encoding. Furthermore, **memory optimisation is approximated by a dependency length analysis**.

The thesis is structured as follows: Chapter 2 gives an overview of translationese research, the interpreting process, and shows what we know about interpretese using mainly traditional corpus-based approaches so far. Furthermore, it introduces *Information Theory* and *Dependency Locality Theory*, and related studies to this work. The chapter ends with the definition of simplification used in this thesis and the research questions investigated in the individual studies. Chapter 3 introduces the EPIC UdS corpus (Przybyl et al., 2022b), an European Parliament interpreting corpus with English and German comparable as well as parallel data, and gives an overview of the methodological design used in the individual studies. The comparable English and German dataset is then first used in an exploratory approach to investigate features of *interpretese* via the information-theoretic measure of relative entropy/Kullback-Leibler divergence (Chapter 4). The results of the exploratory approach are then used to investigate potential effects of optimising the message encoded and focuses on simplification of the interpreting product on different levels: Chapter 5 for lexical investigations, using traditional corpus-based methods and the information-theoretic approach of surprisal. Chapter 6 for a syntactic overview, alignment analysis, and dependency locality analysis. Both chapters take a comparable as well as parallel approach investigating *interpretese* by comparing interpreted language to original source speeches, as well as by looking for explanations of the effects by comparing interpreting to the source speeches where triggers may be found. Chapter 7 combines the results of the individual studies and concludes the findings of this thesis.

Chapter 2

Theoretical background

2.1 Translationese – specific properties of translated language

Translation research has shown that written translations show regularities in their linguistic properties which differ from written texts produced outside the communicative situation of translation. This thesis focusses on simultaneous interpreting as a spoken type of translation. However, spoken translation has not been studied as largely as written translation, and the search for specific linguistic properties of language that interpreters use is rooted in translation studies. Therefore, in a first step, the related key concepts in translation studies are introduced, the specific linguistic properties of written translations are described, and potential explanations for these specific properties are explored.

2.1.1 Key concepts in translationese research

The specific properties of (written) translated language such as specific lexicogrammatical, morpho-syntactic and textual language patterns are generally referred to as *translationese* (Gellerstam, 1986; Santos, 1995), *third language* (Duff, 1981), *third code* (Frawley, 1984), *hybrid text* (Schäffner and Adab, 2001a,b) or even described as universal features of translation (*translation universals*) by Baker (1993).

The term *translationese* as used by Gellerstam (1986) describes these specific patterns as fingerprints or traces of the source language or the translation process that are detectable in the translation product. Two categories of this phenomenon can be distinguished. First, patterns that occur in originals and translations, however, their frequencies differ statistically. Second, patterns that occur only in translated texts, which result in a marked use of the language (Puurtinen, 2003). It is important to stress that the term *translationese* in corpus-based translation studies is generally

not used to refer to poor translation, but rather relates to statistical differences in the use (overuse or underuse) of certain linguistic phenomena that can be found following a comparable corpus analysis: Original texts in one language are compared to translated texts in the same language (Gellerstam, 1986). However, this markedness of translations can sometimes result in 'strange' patterns of linguistic expression or a subuse of the target language (*third code)* (Frawley, 1984).

Occasionally, the term *translationese* is also used as an evaluative term as in Baker (1993) who uses the term to refer to "cases, when an unusual distribution of features is clearly a result of the translator's inexperience or lack of competence in the target language" (Baker, 1993, 249). Furthermore, the term is used in computational linguistic research, where it generally refers to translated texts as such, not to particular linguistic properties (cf. Nikolaev et al., 2020; Carter and Inkpen, 2012, amongst others). This work uses *translationese* and *interpretese* to describe the specific linguistic properties of the translation or interpreting product, respectively, and is not used in an evaluative way.

Some authors even claim that these characteristics of translations are universal to translated texts as such. Baker (1993) claims that the properties of translation reflect the constraints of the translation process. Mona Baker defines universals of translation as "universal features of translations.... which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker, 1993, p.243). Such universal properties of translations may be detected using a comparable corpus study and do not necessarily require a comparison of the translated text to the original text it was based on.

Toury (1980, 2012) sees two main *laws of transnational behaviour*: First, the *law of growing standardisation* is linked to the translator's effort to adapt the translation product to norms in the target language and culture: "in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire" (Toury, 2012, p.304). Second, the *law of interference* can be observed as traces of the source text that are left in the translation product: "in translation, phenomena pertaining to the make-up of the source text tend to force themselves on the translators and be transferred to the target text" (Toury, 2012, p.310). It is therefore seen not to be the translators' choice to opt for a particular linguistic encoding, but rather they are thought to be forced to choose a translation-specific encoding, depending on factors such as experience and translation strategy. These choices are also linked to cognitive processes and socio-cultural conditions in which translation is performed and consumed.

Although later calling the term *universals* "unfortunate" and proposing rather to speak about *tendencies, generalisations, patterns* or *regularities* (Chesterman, 2017, p.252) as it implies that the translation features are universal regardless of text type, cultural context or language pair involved, Chesterman (2004) uses the term to distinguish between features that can be traced back to the source text (S-universals)

and features that can be detected by comparing translations to non-translated texts in the same language (T-universals). S-universals such as *interference, dialect normalisation* or *reduction of repetitions* require corpora of source and target aligned texts (i.e. parallel corpora). The T-universals such as *simplification, untypical pattering* or *under-representation of target language specific items* can be studied with a comparable approach.

Since then, research has rejected the universality claim and instead argued that these features depend on the language pairs and linguistic systems involved (Teich, 2003; House, 2008). The universals hypothesis has been contested for different languages such as English, Finnish, Hungarian, Italian, or Swedish (Mauranen and Kujamäki, 2004; Mauranen, 2007). Regardless of whether these specific properties of translations are universal due to the translation process or rather depend on other factors, it is not disputed that such differences exits. Starting with Baroni and Bernardini (2006), text classification tasks have since been able to successfully distinguish translations from non-translated texts (Baroni and Bernardini, 2006; Volansky et al., 2015; Rubino et al., 2016; Kunilovskaya and Lapshinova-Koltunski, 2020).

In this thesis, specific properties of simultaneous interpreting are investigated, assuming that these features found in the interpreting product reflect the constraints of the interpreting process. As this work is limited only to English and German, it does not intend to challenge the universality claim and rather uses *translationese* and *interpretese* to refer to the linguistic properties defining the use of translated and interpreted language, respectively. This does not rule out the possibility that similar trends may be found in simultaneous interpreting. However, the extent to which the feature occurs may differ (language pair, distance of language pair). Furthermore, the approach taken in this thesis is a comparable study to describe specific characteristics of interpreted language, as well as a parallel approach to look into potential explanations of characteristics found in interpreting.

2.1.2 Simplification in the context of other translationese features

When it comes to describing characteristics of translated language, Baker finds that translations tend to show properties such as *simplification, explicitation, normalisation / conservatism* and *levelling out* (Baker, 1996, p.176-177). She uses a comparable approach, comparing translations to original texts in the same language and bases her study mainly on the language pair English-Arabic, but also considers findings of related studies to validate these properties. Other authors add *shining-through* (Teich, 2003) or *interference* (Toury, 1995) to the specific characteristics of translated language, which requires a parallel investigation of translations and the sources texts on which they were based.

As mentioned above, *simplification* is generally investigated via a comparable corpus approach, comparing an original, non-translated text to a translation in the same

language as well as of the same domain, text genre, time frame, etc. It is a subconscious tendency to simplify the language and/or message of the source (Baker, 1996, p.176). This definition is very broad and expands to many aspects of simplification. It may include all linguistic levels, in particular lexis and syntax, but also semantics of the message. Regarding lexical aspects, Blum and Levenston (1978) perceive lexical simplification as "the process and/or result of making do with less words" (Blum and Levenston, 1978, p.399).

This definition points towards reduced lexical variation in translations. With this focus, Laviosa (1998) shows that English translations use a narrower range of vocabulary than comparable originals. She uses lower lexical density and greater percentage of high-frequency words in translations compared to non-translations to make her claim. Furthermore, she shows that the most frequent words in a corpus (i.e. list heads containing 108 words) are repeated more often in translations than in non-translations and finds that the translation's list heads also contain fewer lemmas compared to non-translations. (Laviosa, 1998, p.563). In line with these findings are multiple studies on lexical variation, showing that lexical variation (e.g. as indicated by a lower type-token ratio (TTR)) is generally lower in translations compared to originals in the same language (Sheikh Al-Shabab, 1996). However, these tendencies seem to depend on the language pairs and register (Hansen-Schirra, 2011; Steiner, 2013).

Volansky et al. (2015) also use a comparable corpus approach and compare English original texts from EUROPARL (Koehn, 2005) to translations into English from 10 different source languages. They use various features of lexical simplification in a machine learning text classification task. The authors find that *lexical density* measured as the frequency of tokens that are not nouns, adjectives, adverbs, or verbs compared to all tokens is not suitable to correctly predict whether a text is an original or a translation. The classification results are close to chance level (accuracy 53%). On the other hand, *lexical variety* measured as type-token ratio (TTR) (translations are assumed to use less divers vocabulary), indirectly measured as *mean word length* (translations are predicted to use simpler and therefore shorter words) and *syllable ratio* (translations are assumed to use words with fewer syllables) were found to be reliable predictors for detecting translation. Prediction accuracy for these features range from 61-76%. *Mean word rank* calculated via a list of 6000 most frequent words in English, and their position in the frequency ordered list was also found to be a good predictor of translations. Translations are found to use more frequent words more often (prediction accuracy in the classification task is 69 and 77%). *Most frequent words* as the last lexical feature of simplification used in Volansky et al.'s (2015) classification task is closely related to Laviosa's (1998) study, giving a correct prediction in the classification task for 64% of the texts (Volansky et al., 2015). Halverson (2003) gives a theoretical explanation for such findings, saying that translators use more prototypical language, and Shlesinger (1989) states that translators tend to go toward the mean.

From a syntactic perspective, simplifying can mean breaking up long sentences into shorter ones during the translation process (Baker, 1996, p.181). Although strictly speaking, such an analysis would require a parallel approach, comparing translations to the source speeches they were based on, such patterns may also be the result of the translation process with shorter sentences in translations observed when compared monolingually to original texts in the same language. Generally, mean sentence length is found to be shorter in translations than in comparable originals (Baker, 1996; Laviosa, 1998; Neumann, 2003; Corpas Pastor et al., 2008). In addition to shortened sentences, Toury (1995) adds a simpler sentence structure as an indicator of syntactic simplification. These shallow syntactic features are mostly investigated by a comparable approach, as the source and target may generally not be directly compared due to morphological and syntactic differences in the linguistic systems, e.g. use of compounds (counted as one word in German) vs. several words in English (Lapshinova-Koltunski, 2015).

In their text classification task mentioned above, Volansky et al. (2015) also use a syntactic predictor of translations: *mean sentence length* gives good classification results with 65% prediction accuracy, however, opposite to the expected trend. They find that the English translations are generally longer than English comparable originals. Moreover, they find that this trend depends largely on the source language the translation was based on. This contrasts to the previously observed trend of translations using shorter sentences than comparable originals (cf. Baker, 1996; Laviosa, 1998; Neumann, 2003; Corpas Pastor et al., 2008) and shows the influence of the language pair involved in the translation. Other studies on syntax (Eskola, 2004a), as well as lexis and syntax (Eskola, 2004b) show that the expected syntactic simplification in the translation product is not always confirmed.

Providing an argument for the existence of simplification as a general tendency in translation, Ilisei et al. (2010) use a machine learning approach to detect *translationese*. They test whether adding simplification features to a text classification setup significantly improves classification results in a comparable study of Spanish originals and translations into Spanish. In line with the studies mentioned above, adding lexical simplifications features (*word ambiguity* – average number of senses per word derived from Spanish Wordnet; *word length* – syllables per word; *lexical richness* – TTR; and *information load* – proportion of lexical words to all tokens) improve the classification results. Moreover, they also use syntactic simplification features in their text classification experiment: *average sentence length*, *sentence depth* (measured as parse tree depth), and *sentence complexity ratio* (the proportion of simple sentences, complex sentences, and sentences without finite verb). From the syntactic feature set, it is mainly the feature *sentence length* that helps the classifier distinguish between originals and translations. Using all combined simplification features helps to differentiate between translated and original texts with 97% accuracy, and therefore Ilisei et al. conclude that simplification does exist in translated texts. However, *sentence depth* and the proportion of simple to complex sentences

do not seem to contribute much to the improvement in classification. Moreover, Corpas Pastor (2008) find translations to show less syntactic simplification with respect to complex sentences and depth of syntactic trees in a study on Spanish medical and technical translations compared to Spanish originals. In a related study also in the medical and technical domain, Corpas Pastor et al. (2008) show that translations have a smaller proportion of simple sentences, however, use significantly shorter sentences than non-translations. Furthermore, they show that translators' experience also influences simplification tendencies. More traces of simplification were found in translations by professionals vs. non-professional translators.

From the studies discussed in this section, it can therefore be summarised that overall lexical simplification seems to be very prominent, whereas concerning syntactic simplification other factors such as language combination, text domain or professional status may influence the simplification tendencies.

Generally, the result of simplification is "making things easier for the reader" (Baker, 1996, p.182). However, some trends found in translations may counteract simplification tendencies in the translation product. Translations are also found to avoid repetitions (Schlesinger, 1991; Toury, 1991), which may lead to shorter target texts on the one hand, but also to more implicit texts that are not necessarily easier to understand for the reader. *Implicitation* as defined by Klaudy and Károly (2005) is "when a source language (SL) unit with a specific meaning is replaced by a target language (TL) unit with a more general meaning; when translators combine the meanings of several SL words in one TL word; when meaningful lexical elements of the SL text are dropped in the TL text; when two or more sentences in the source text (ST) are conjoined into one sentence in the target text (TT); or, when ST clauses are reduced to phrases in the TT, etc" (Klaudy and Károly, 2005, p.15). This definition aims at a parallel comparison of source and target and shows that *simplification* and *implicitation* may be seen as counteracting tendencies. Leaving things implicit rather than spelling them out renders the target text more vague and therefore more difficult for the reader to understand, i.e. less simplified. Joining two or more source sentences into one target sentence without omitting the source text content would make the syntactic structure more complex rather than simplified.

Explicitation as an opposite tendency to *implicitation* can be seen rather as a trend in line with simplification. In translations, it is a tendency to spell things out rather than leave them implicit (Baker, 1996), or, opposite to Klaudy and Károly (2005)'s definition of *implicitation*, *explicitation* can be observed "when a SL unit with a more general meaning is replaced by a TL unit with a more specific meaning; when the meaning of a SL unit is distributed over several units in the TL; when new meaningful elements appear in the TL text; when one sentence in the ST is divided into two or several sentences in the TT; or, when SL phrases are extended or "raised" to clause level in the TT, etc" (Klaudy and Károly, 2005, p.15). Replacing a general lexical word of the source with a more specific word in the target as a sign of

explication overlaps with Toury's understanding of *simplification* (Toury, 1995) which links less ambiguous expressions to simplified target texts. Furthermore, splitting a long complex sentence into several shorter less complex sentences has also already been covered by syntactic simplification as described above.

Blum-Kulka (1986) first found *explicitation* as a tendency in translation to use more redundancy and explicit cohesion than the source texts on which they were based. Translations generally use more explanatory lexis such as *cause, due to, lead to, etc.* and connectors, more syntactic repetition, and textual extensions (*parentheses, dashes, footnotes and paraphrasing of metaphors*) (Baker, 1996). All these efforts to make the text more explicit can also be seen as making the text easier for the reader, or as Kuo (1993) sees simplification: simplified texts generally use increased redundancy and amplified explanation (Kuo, 1993 in Crossley et al., 2007).

Furthermore, *interference* (Toury, 1995) as mentioned above or *shining trough* (Teich, 2003), i.e. when the translated text is more orientated towards the source text and the source language shines through, may lead to *untypical lexical patterning* or *unique items* (Chesterman, 2004). It is important to note that not all of the traces of the source texts are specific patterns that occur only in translations. Sometimes these traces of the source may result in atypical frequencies of lexico-grammatical patterns that also exist in the target language; however, they are used more or less frequently in translation than comparable original texts of that language. Examples for such source language patterns shining-through can be more frequent use of passive voice in English translations as a shining-through effect from German source texts (Teich, 2003) or deviating frequencies for pre- and postmodification in English translation based on German sources (Teich, 2003; Hansen-Schirra, 2011). Overall, these traces of the source texts may lead to unexpected patterns in the target texts and thus less simplified texts.

In the context of *shining-through*, a further tendency in translation must be mentioned. Instead of leaving traces of the source text that can be observed in the target text, translations are also found to show an opposite trend and show traces of *standardisation* (Toury, 1995). This tendency is sometimes also called *conservatism* or *normalisation* and can be described as a tendency to conform to typical patterns of the target language (Baker, 1996). It means "that translations are supposed to avoid margins or periphery and remain safely within the mainstream" (Mauranen, 2007, p.12). Here too, an overlap with *simplification* can be found. *Normalisation* can also lead to high frequencies of certain lexical patterns, use of common lexis instead of unusual lexis, preference of standard language over dialect, etc, which may lead to a simpler target text. This has been supported by a greater proportion of n-grams or multi-word clusters (Nevalainen, 2005 in Mauranen, 2007).

A clear distinction of *simplification* or other translationese tendencies seems to be difficult, and the definitions given above for *explicitation, implicitation*, and *normalisation* overlap with the idea of *simplification* of "making things easier for the reader" (Baker, 1996, p.182). Considering all these tendencies observed in translations that

can be considered to counteract or reinforce simplification, it is not surprising that Mauranen (2007) states that "the simplification hypothesis has not been generally supported or refuted" (Mauranen, 2007, p.11). As discussed above, the results for syntactic simplification show some mixed trends, depending on the languages and text domain involved in the studies.

2.1.3 Translationese and potential explanations

As established above, translations are known to differ from original production in comparable settings (same register, time frame, etc.). Furthermore, it was shown that translation scholars try to combine these differences into feature groups, however, sometimes with overlapping or contradictory categories. As to potential causes for these specific patterns in translation, the literature also gives a multitude of reasons. Mauranen (2007) attributes these differences to tendencies "which can be traced back to general human cognitive capacities and those which relate linguistic structures and the functional uses of language" (Mauranen, 2007, p.6). The latter is acknowledged by scholars who see the influence of the source language, but also cultural differences such as cross-linguistic divergence between registers in the source and target language as a main cause of linguistic differences (e.g. Gellerstam, 1986; Santos, 1995; Rabadán et al., 2009). There may be lexical voids in any given language pair, with one language lacking a term that exists in the other language. Avoidance of certain lexical items (due to appropriateness in different contexts) or ever the choice of translators or editors to favour certain linguistic expressions over others (Blum and Levenston, 1978). This may also be related to an observed 'language prestige', and unequal status of some languages and cultures that can have an effect on interference being tolerated in one target language but not another. Mauranen and Kujamäki (2004) show, for example, that the Finnish culture is more open to interference from English than from Russian (Mauranen and Kujamäki, 2004). In line with these findings, Evert and Neumann (2017) find that German shows more traces of the English source texts than the opposite translation direction (Evert and Neumann, 2017).

Taking this approach to explain the linguistic characteristics of *translationese* would mean that we could expect the same differences when comparing interpreting and spoken originals, i.e. *interpretese*. However, the first potential cause for the linguistic differences between translations and originals, as stated by Mauranen (2007) relates to the general human cognitive capacities and refers to the constraints of the translation process as such (Baker, 1993). This approach is supported by studies that show that translationese tendencies like simplification also depend on qualities of the translator itself. Translator experience has a significant influence on the simplification properties of the translation product, with professional translators simplifying more than non-professionals (Corpas Pastor et al., 2008). Furthermore, translators may sometimes intentionally adapt their target texts to a target audience. Such audience design can, for example, be seen in an additional explanation of cultural items (case

of *explicitation*) or avoidance of repetitions (case of *implicitation*).

The written translation process in which translators transfer the meaning of the source text to the target text, sometimes in a non-linear way as they may jump back and make immediate corrections (cf. studies on eye-tracking, e.g. Schaeffer et al., 2019b), including multiple revision rounds by the translator themselves but also by external revisers or editors, is very different from the mostly linear translation process in simultaneous interpreting. Although the interpreting process is explained in more detail in the following section (Section 2.2.1), the most obvious difference is the increased time constraints and the very limited possibilities of making changes to the spoken translation product. Therefore, it can be assumed that the process-related features of *translationese* will differ greatly from the features found in *interpretese*.

On a more general note, some methodological issues in the quest for translationese features in general and simplification in particular have to be stated. Although most of the studies mentioned above employ a comparable corpus analysis and compare translations to original texts in the same language, the definitions of translationese features sometimes mix the comparable perspective with explanations that are linked to the source texts (parallel analysis). For example Baker (1996), although stating that a comparable perspective is sufficient to study specific properties of translations, gives explanations for the observed result of shorter sentences in translations, such as breaking up long sentences during the translations process into shorter ones. In order to validate such explanations, a parallel corpus analysis is required to compare translations with the sources they were based on. In order to resolve this methodological shortcoming, this thesis uses both perspectives: a comparable corpus approach to see how interpreted texts differ from non-mediated spoken texts, but also a parallel corpus analysis, where the source texts (original speeches) are investigated to find explanations and potential triggers for specific phenomena that are found in the target texts (interpreted speeches).

Furthermore, translationese features that are generally employed in comparable translationese studies, features such as type-token ratio, lexical density, indicators lexical variation, sentence length, POS n-grams and others (e.g. Volansky et al., 2015; Corpas Pastor et al., 2008), are easily extractable features that are able to capture linguistic differences of translated and non-translated texts. However, they do not reflect the cognitive processes that underlie language perception and production. Therefore, it may be useful to apply corpus-based measures that are known to provide a direct link to cognition, such as *surprisal* (Hale, 2001) or *dependency locality measures* (Gibson, 2000). Both measures are linked directly to linguistic processing and therefore may shed light on the cognitive side of the translation process, captured in the translation product. The measure of *surprisal* touches aspects of lexical variation as measured in the traditional corpus-based approaches, as it measures how predictable a linguistic item is in a given (variable) context. As for syntactic simplification, some studies mentioned above use measures of dependency grammars (e.g. Corpas Pastor et al., 2008), however, rather as an overall measure of *sentence depth*

measured as parse tree depth, and not as the dependency distance of individual syntactically related words in a sentence to approximate the effort required to maintain a linguistic item in working memory. Using these more process-related measures is hoped to give a more comprehensive picture of process-related traces that can be observed in the translations product.

And last, the definition of simplification as laid out above is somewhat fuzzy and overlaps with other translationese tendencies. This thesis therefore employs a wider definition of simplification, capturing the general notion of "making things easier for the reader" and potentially also for the producer, which may include explicitation and normalisation tendencies. A detailed definition of simplification as used in this thesis is given in Section 2.5.2.

2.2 Simultaneous interpreting and cognitive constraints

As laid out in the previous section, at least some of the causes of *translationese* and potentially also *interpretese* are process related. The translation process for simultaneous interpreting as a spoken mode of translation differs significantly from the translation process for written translation. Although certain time constraints also exist for written translation, these constraints are more imminent in spoken translations and simultaneous interpreting, in particular. The next section first explains the interpreting process and particular constraints, using various theoretical interpreting models. Then it shows ways of potentially coping with cognitive constraints in simultaneous interpreting via strategies that interpreters may employ, which potentially may have influence on particular linguistic characteristics found in the interpreting product. At the end of the section, ways to measure cognitive load in simultaneous interpreting are shown, again stressing the importance of cognitive aspects in simultaneous interpreting research. It is important to note that this refers only to simultaneous interpreting and does not cover other spoken translation modes such as consecutive interpreting, dialogue interpreting, or others.

2.2.1 Interpreting process and models

Simultaneous interpreting research acknowledges the particular role of cognitive load as a result of immediate time constraints during the translation process in simultaneous interpreting. Many theoretical interpreting models try to capture the interpreting process and how limited cognitive resources are managed (Gerver, 1976; Moser, 1978; Setton, 1999; Gile, 2009, 2017; Camayd-Freixas, 2011; Seeber, 2011, 2013, amongst others). Although, they emphasise different aspects of the interpreting process, they all see particular implications for interpreter's (short-term) memory and high cognitive demand.

One of the most widely used frameworks to explain cognitive processes in simultaneous interpreting is Gile's *Effort Models* (Gile, 1995, 1997, 2009, 2017). As these Models were meant as an explanatory concept for interpreting students, they give a good overview about general processes that take place when interpreting simultaneously. Moreover, they are also often used in theoretical and empirical research; however, the Effort Models are also criticised for different shortcomings. In the following, Gile's *Effort Models* are described in order to give an overview of the interpreting process and related load on the cognitive system.

According to Gile (1995), simultaneous interpreting can be defined as the sum of a comprehension oriented "Listening and Analysis Effort (L)", which includes processes from perceiving the incoming sound wave to comprehending the meaning of an utterance, a "Production Effort (P)", including a mental representation of the message, speech planning and production, and a "Memory Effort (M)" for short-term memory when original speech segments have to be stored intermediately until they are produced as target segments. Later, a "Coordination Effort (C)" was added to the model, acknowledging the importance of attention allocation between the other three efforts (Gile, 1997) (cf. Formula 2.2.1).

$$SI = L + M + P + C \tag{2.1}$$

It is important to note that the mathematical signs "=" and "+" are not to be used in the pure arithmetic sense. The sum of L + P + M + C cannot exceed the Total Available Processing Capacity (TAPC) and the capacity available for each effort must be equal to or higher than the required effort (Gile, 1997, p.199). All these processes take place simultaneously and within a very short time frame: ear-voice span (EVS), i.e. the time span between L and P was found to be mostly between 1 and 3 seconds (Defrancq, 2015).

Gile's Effort Models assume that resources can be shifted between and reallocated among tasks. Processing capacity supply and demand are constantly competing with each other. The processes of listening and analysing, producing the target text, and intermediately storing content in memory are seen as non-automatic tasks, and therefore these three efforts compete for available processing capacity (Gile, 1997). Relying on evidence from psychology and psycholinguistic research, Gile assumes that each process can be (partially) controlled and that these three processes are less automatic in simultaneous interpreting than in other communication situations, thus requiring more efforts in simultaneous interpreting. Gile also argues that comprehension in simultaneous interpreting is a more active process than in other situations of speech comprehension, as interpreters need to actively listen all the time, whereas other conference participants can practice selective listening and only focus on what is interesting for them. Furthermore, the Listening and Analysing Effort (L) can be seen to be more intense in interpreting, as Gile expects interpreters to be less familiar with the subject matter and terminology than the conference participants to

whom the source speech is addressed (Gile, 2009). However, it can be argued that in some conference interpreting settings, such as the plenary sessions of the European Parliament, on which the studies in this thesis are based, at least some of the permanent interpreters may be as experienced and involved in the matters discussed as the delegates. The Production Effort (P) is also seen to be more effortful than under regular speech production conditions, as the interpreter has to render somebody else's speech, and therefore cannot use the strategies for improving fluency they would use in conventional speech (Gile, 2009). It can be assumed that interpreters have their own mechanisms, such as the use of increased formulaicity (Kajzer-Wietrzny and Grabowski, 2021) or inserting fillers such as filled pauses (Defrancq, 2018) as means to improve the fluency of their output. It can be argued to which extent such mechanisms may be at least semi-automatic. Furthermore, the means of simplification that are investigated in this thesis, such as producing expected words in context (i.e. low surprisal), could also be seen as strategies to improve the fluency of the interpreted speech.

Gile (1999) hypothesises that individual processes compete for resource allocation and that the total capacity of resources cannot be exceeded. Simply speaking: if the Effort for Listing and Analysing requires more resources because of informationally dense or high surprisal input speech, unfamiliarity with the topic of the speech, an unknown accent of the source speaker, or other reasons, other Efforts cannot be allocated enough resources, and, for example, Memory or Production Effort may suffer.

In addition to explaining the interpreting process in a very simple way, Gile's *Effort Models* make an important claim about interpreters' cognitive processing. As stated in the *tightrope hypothesis*, Gile (1999) assumes that interpreters work close to cognitive saturation most of the time. Increases in processing-capacity requirements or mismanagement of cognitive resources can result in deteriorating interpreter's output. Gile (2017) argues that if interpreters generally work well below cognitive saturation levels, errors and omissions would only occur when significant difficulties arise in the source speech. This thesis investigates simplification as a general means of efficiently managing cognitive resources and staying well below the level of cognitive saturation.

Gile (1995) also acknowledges sources of error, omissions, and infelicities. Besides cognitive overload, insufficient linguistic or extralinguistic knowledge can also lead to errors and omissions in interpreting (Gile, 2009). Furthermore, the author also takes into account external factors, as well as speech-internal problem triggers, that are grouped according to processing capacity requirements. As problem triggers associated with a required increase in processing capacity, Gile (1995) lists high density of source speech which might come with high rate of delivery or high density of information content. The latter can directly be linked to the approach taken in this thesis to investigate unexpected words in context, i.e. high surprisal words. The interpreter is required to process more information per unit of time in such conditions, requiring more Listening and Analysis Effort as well as Production Effort (Gile, 1995, p.172 ff.).

In conference interpreting settings such as the European Parliament plenary sittings, some of the speeches are prepared and read-out. It can be expected that they are more informationally dense and delivered faster as the speaker tends to avoid hesitations, as well as filled and unfilled pauses (Halliday, 1989). Furthermore, external factors such as poor quality of sound, strong accent, incorrect grammar, or unfamiliar lexical use can increase efforts required for listening and analysing, as well as unusual linguistic style and reasoning style (Gile, 1995, p.173). Unfamiliar technical terms and unknown long names (with high information density) not only increase the processing capacity requirements for the Listening and Analysis Effort, but also increase the capacity requirements of the Memory Effort. Saturation of the Memory Effort can also occur if the source and target languages are syntactically very different and the source speech input needs to be stored intermediately in memory before it can be produced in the correct syntactic position in the target speech (Gile, 1995, p.174).

The problem triggers mentioned above are associated with an increase in processing capacity requirements. Other problems are associated with signal vulnerability. Gile states that numbers, short names and acronyms do not necessarily require much processing capacity, but are particularly vulnerable to "momentary processing capacity shortage", as they are short in duration and low in redundancy (Gile, 1995, p.174).

It is important to note that problem triggers alone do not necessarily lead to problems with processing capacity. However, they can be seen as potential sources of errors and omissions. The larger context needs to be considered to assess whether a problem trigger leads to a problem in the output. For example, an informationally dense segment that occurs at the end of a sentence and is followed by a pause would no longer require Listening, and Analysis Effort during the pause, and therefore all processing capacity can be used for Memory and Production Effort (Gile, 1995, p.174).

In addition to these Efforts competing for momentary processing capacity, Gile (2008) also refers to the concept of *imported load* or *knock-on effects* in simultaneous interpreting. Source text related difficulties may lead to *knock-on effect* or *failure sequences* at a later point in the interpreting product. According to Gile (2008), effect of *imported load* may even occur as far away as at the end of the next sentence. However, empirical evidence does not support this claim (Plevoets and Defrancq, 2020; Chmiel et al., 2023).

Gile's Effort Models include many aspects that have to be considered within the simultaneous interpreting process. The Models illustrate simultaneous interpreting as complex task in schematic simplicity, and at the same time provide a holistic view of the interpreting process. One main criticism of Gile's Models is that he sees the main source of errors and omissions in the interpreting product as mismanagement of cognitive resources. Other authors also see a vital role in the distribution of attention to the individual subtasks in interpreting (Gerver, 1976; Moser, 1978). However, Gile assumes that interpreters can freely allocate the available resources to whatever Effort requires attention; he ignores the fact that not all the same efforts potentially tap into

the same pool of processing effort. In fact, experimental research on working memory management shows that there may not only be one general pool of cognitive resources, as the workload of a primary task does not necessarily affect the performance of a secondary task (D. Alan Allport and Reynolds, 1972; Wickens, 1976). Structurally similar tasks might interfere more with each other than tasks that are structurally dissimilar (Wickens 2002 in Seeber, 2013).

Furthermore, it can be questioned whether interpreters really work close to cognitive saturation at all times (tightrope hypothesis). Gile bases this assumption on the high number of errors and omissions in the interpreter's output. If the tightrope hypothesis is true, it can be assumed that interpreters constantly try to keep the required efforts low. As they cannot influence the Listening and Analysis Effort during the interpreting process, they may try to maintain Memory Effort low by producing source text items as early as possible and by keeping dependency distances between individual words in the target output as small as possible. Furthermore, the Production Effort can also be influenced by producing output that requires the least amount of effort, i.e. the use of formulaic expressions and expected words in context (low surprisal words). Although the Effort Models were not designed to be tested empirically, the approach taken in this thesis – to investigate if interpreters generally aim to simplify their output – can be seen in the context of the tightrope hypothesis.

The importance of limited cognitive resources that interpreters have to manage efficiently when working in the simultaneous mode is also acknowledged by Seeber's (2011, 2013) Cognitive Load Model (CLM) - with cognitive load even being part of the model's name. Seeber's CLM was generally based on the Effort Models (Gile, 1995, 1997, 2009) and also aims at capturing and illustrating the cognitive demands inherent in simultaneous interpreting (Seeber, 2011). However, Seeber rejects the claim that interpreters work close to cognitive saturation at all times. Moreover, Seeber assumes that, rather than drawing from only one general pool of undifferentiated cognitive resources where these resources can be freely shifted and re-allocated among tasks, there is a conflict potential caused by overlapping and interfering tasks (Seeber, 2011, p.189). In line with Wicken's (1984 in Seeber, 2011) Multiple Resource Model, Seeber bases his theory on the assumption that when two or more individual tasks are combined (such as a language comprehension and language production task, combined with output monitoring and intermediate storing of information), the sum of the individual tasks requires more processing capacity than the sum of the individual subtasks conducted alone. Such task interference is assumed to be stronger when the tasks are structurally similar.

The author aims to describe the cognitive processes and workload involved during simultaneous interpreting as a specific language processing task. He uses a conflict matrix to show interfering and overlapping of concurrent tasks by summing the load for storage, perceptual auditory verbal processing, cognitive verbal processing, and verbal response processing to indicate the total load. One of the strengths of the

model is that Seeber differentiates cognitive effort during different interpreting conditions: interpreting symmetrical or asymmetrical structures, involving strategies such as waiting, stalling, chunking, and anticipating. The assumed cognitive load varies, depending on the micro strategy used by the interpreter and, if the interpreter is not able to follow the syntactic structure of the source input in their target output (asymmetric condition), the cognitive load is assumed to be considerably higher than for producing symmetrical output (cf. Figure 2.1). The CLM tries to quantify cognitive load and was set out to be empirically applicable; however, it still awaits to be tested on a larger scale. The strength of the model is the inclusion of different interpreting conditions and assumed quantitative effects on cognitive load, although it is not clear how to operationalise this on a larger scale. One weakness of the model is also acknowledged by the author himself. The CLM ignores factors that influence retrieval and activation mechanisms that may influence cognitive load, such as semantic priming, relative word frequency, abstractness and concreteness, or cognate status of source and target words (Seeber, 2011, p.192).

In this context, it is important to take a closer look at how interpreters deal with different types of source input and the strategies they may use. As they are not able to influence the source input, including the organisation of propositional information, professional interpreters may use different macro-strategies depending on the input they have to deal with. These may be particular behavioural patterns that interpreters have acquired during training and interpreting practise as an attempt to deal with the constraints they face during the interpreting task (Riccardi, 1998). It can be assumed that interpreters may opt for a strategy that is beneficial for cognitive processing demands (Seeber, 2011). For example, when it is possible to follow the word order of the source in the interpreter's target output, the interpreter is assumed to maintain the source word order (first-in-first-out processing) and opt for the least demanding strategy (Shlesinger, 2000). Looking at an individual proposition, Seeber (2011) assumes that once production begins, the overall cognitive demands remain fairly constant for such symmetrical structures and decrease slightly after comprehension of the input sentence is completed (cf. Figure 2.1, syntactic symmetry, baseline condition). It is important to note that such structures can be interpreted with relatively short lag, which means that items do not have to be stored in working memory for long, but also that the interpreter can focus on the next proposition uttered by the source speaker and reduces the risk of interference between two tasks (Wickens 1984, 2002 in Seeber, 2011). However, this strategy only works for language pairs that are syntactically symmetrical, and do not require major repositioning of syntactic elements. The CLM (Seeber, 2011) also gives possible interpreting strategies that can be used for syntactically asymmetrical language pairs, and shows their assumed effect on cognitive load. The interpreter may *wait* for more input. This results in having to store more information in memory until it can be encoded in the interpreter's output. This strategy "allows interpreters to alleviate cognitive load temporarily, (...) trans-

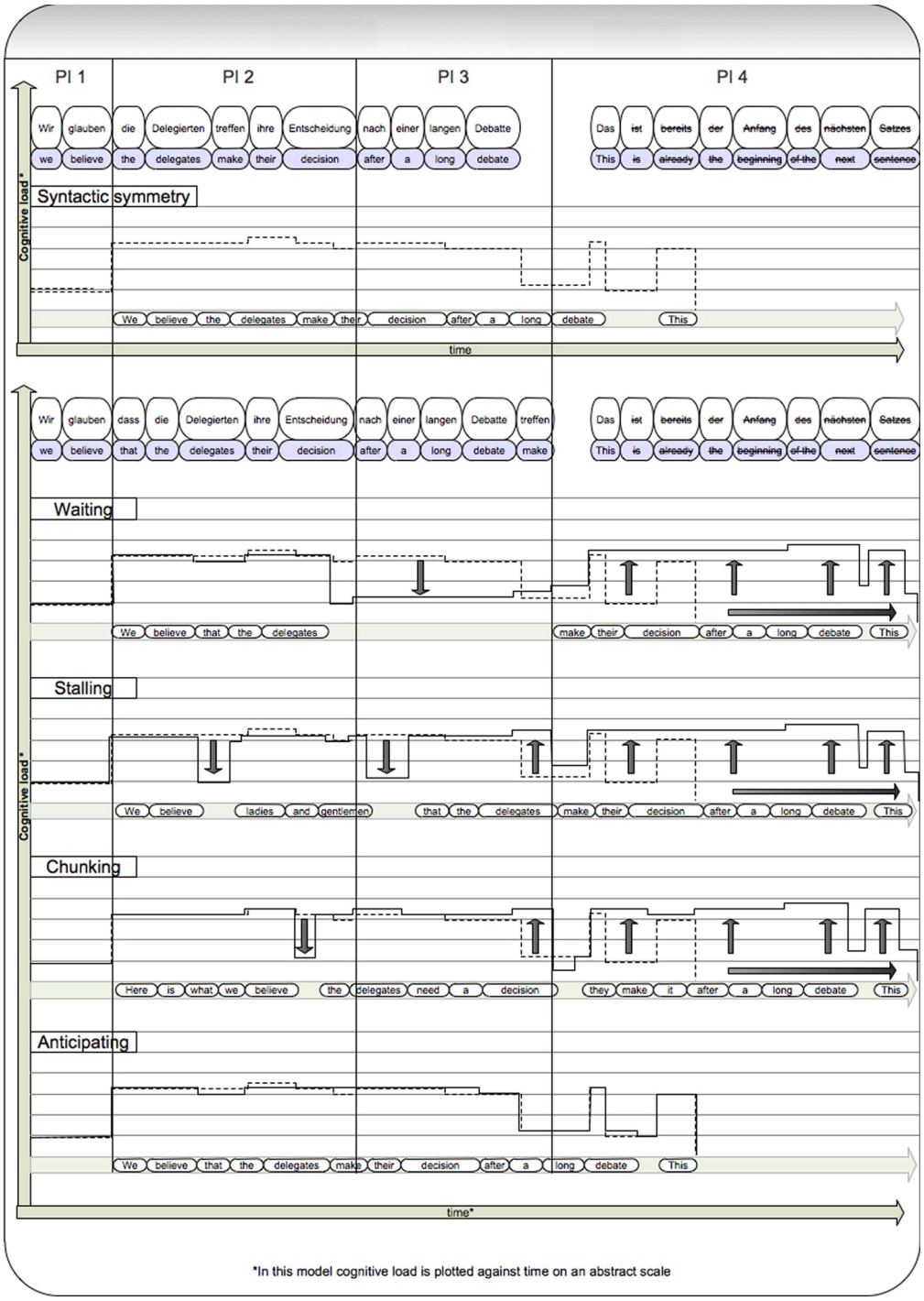


Figure 2.1: CLM for SI. Assumed demands on cognitive resources against time on abstract scale. German verb-initial structure interpreted into English (baseline) vs. verb-final (strategies: waiting, stalling, chunking, anticipating). Vertical arrows: variation of local cognitive load vs. baseline. Horizontal arrows: time lags vs. baseline (Seeber and Kerzel, 2012a, p.230).

forms the process into a simple comprehension and memorisation task. On the other hand, it might cause a spillover effect, leading to a considerable increase in cognitive load downstream" (Seeber, 2011, p.193).

Similarly to *waiting*, the purpose of *stalling* (cf. Figure 2.1) is also to buy the interpreters more time. In contrast to *waiting*, this is not achieved by a (silent) pause, as pauses can result in an uncomfortable expectation within the listeners. Instead, the interpreter fills the time they wait for the missing source speech item by filling the time without adding new information (Seeber, 2011). Such "neural padding" may be repetitions of previously produced speech components or adding neutral fillers (e.g. Ladies and Gentlemen), although strictly speaking *stalling* may be seen as a slight alteration of the source message as it can change the focus or stress of the original message. More importantly, it adds additional processing complexity to the overall task, as encoding and production of padding material overlap with the ongoing comprehension processes (Seeber, 2011).

Chunking (cf. Figure 2.1) refers to the strategy where interpreters segment a larger input unit into smaller fragments to be encoded without having to wait for the entire sentence to be completed. Although the interpreter can start encoding their message as soon as possible, without knowing how the input ends, the earlier produced chunk may have to be strung together later on in order to establish the original meaning. This may result in a temporal increase in cognitive load (Seeber, 2011, p.194). However, compared to other strategies such as reordering, chunking seems to be less effortful (Ma et al., 2020) and reduces syntactic complexity.

The strategies *waiting, stalling*, and *chunking* all result in a substantial lag (increased EVS), which may lead to higher storage costs further down the interpreting process. Seeber (2011) and Gile (1999) believe that this leads to potential spillover effects and potential problems or even failure of the process.

The last strategy described in the CLM is *anticipation*. It is seen as a non-automatic strategy where the interpreter tries to guess a sentence constituent such as the verb before it was uttered in the original. As the interpreter does not wait for the constituent to be uttered, this strategy does not increase the time lag, but according to Seeber (2011) temporarily increases the local cognitive load when the missing constituent is inferred. Generally, this strategy results in cognitive load levels close to the baseline setting. Although *anticipation* is a widely discussed phenomenon in interpreting research literature (cf. reference list in Seeber, 2011), corpus data show that it is a fairly rare strategy used by interpreters, with only 1.5% of time-aligned corpus data showing a time lag of less than one second (Defrancq, 2015, p.31). However, this may depend on the languages involved in the interpreting process with an SOV language interpreted into an SVO language using more anticipation than SVO to SVO (Donato, 2003). Interpreting between structurally similar languages tends to use *anticipation* only very infrequently (Bartłomiejczyk, 2006).

To summarise, the different interpreting models vary in the amount of detail used

to describe the different subprocesses of simultaneous interpreting, strategies that are applied, or the theoretical conclusions drawn: whether interpreters work close to cognitive saturation most of the time (Gile, 1999) or below saturation level part of the time (Seeber, 2011). The two models discussed in the section, as well as other interpreting models in the literature (cf. Ahrens (2024) for an overview) all have in common the acknowledgement that simultaneous interpreting is a cognitively demanding process which presumably has effects on the interpreting product. This thesis does not aim to challenge existing interpreting models, but rather tests the effects of this cognitively highly demanding process via traces of simplification in the product data.

2.2.2 Interpreting strategies

Seeber's (2011) CLM gives a good overview of strategies interpreters may apply when dealing with syntactic asymmetry. However, there are other potential difficulties, including context, terminology, names, speed of delivery, high information density, etc., that interpreters may have to face. Assuming that such strategies employed during the interpreting process leave traces in the interpreting output and in order to know how to interpret potential patterns that may be observed in the interpreting corpus data used in this thesis, the present section gives an overview of potential strategies interpreters can resort to in order to deal with the complex cognitive process of simultaneous interpreting. Interpreting strategies receive a lot of attention in interpreting research (Gile, 1995; Kohn and Kalina, 1996; Kade, 1967; Kirchhoff, 1976; Riccardi, 1998, 2005; Zanetti, 1999; Bartłomiejczyk, 2006; Liontou, 2011, amongst others), although they are sometimes also controversially discussed as to whether they are a conscious choice of interpreters to deal with the constraints they face or rather a by-product of language mediation (Gumul, 2006).

Generally, strategies in interpreting are seen as specific procedures to prevent or solve problems when producing an adequate and coherent interpretation. Interpreters choose to opt for a strategic procedure. This does not just refer to one strategy, but rather a combination of individual subsequent strategies that may depend on each other, trigger each other, or overlap with each other. Interpreters may use these strategies consciously or unconsciously (Wörrlein, 2007).

Different scholars try to categorise the strategies used by interpreters from various perspectives. Wörrlein (2007) differentiates between *micro strategies* that are used at the lexical level and *macro strategies* that are applied across sentences or the entire speech. Christoffels and de Groot (2005) distinguish between *meaning-based strategies*, involving full conceptual comprehension, and *transcoding* where processing is done no deeper than needed, also allowing shortcuts at discrete levels of language processing. Riccardi (2005) categorises *knowledge-based and skilled-based strategies*, Bartłomiejczyk (2006) refers to *problem-oriented and overall strategies*. Al-Khanji et al. (2000) distinguish between achievement strategies to deal with a given problem

and reduction strategies to avoid communicative problems. As described above, Seeber (2011) differentiates between strategies for syntactic symmetry and asymmetry. Kalina (1998) groups the strategies into *comprehension enhancing strategies* and *target text production strategies*. As Kalina's framework is one of the most comprehensive catalogues of interpreting strategies (Liontou, 2011), her categorisation is used here as a basis for an overview of possible strategies that interpreters can use in simultaneous interpreting.

Kalina (1998) refers to *comprehension enhancing strategies* as processes that generally also occur in monolingual communication; however, in interpreting they have to be adapted to interpreting processing capacities and efficiency (Kalina, 1998, p.115). The most general strategies used by interpreters are *preparation strategies* in which the interpreter aims to compensate for a knowledge deficit they may have compared to other conference participants. This includes gathering information on the conference topics as well as individual speeches, including speaker manuscripts and presentation slides, but also information on the participants such as their names, titles, background, etc. This mainly includes preparing terminology in the source and target languages that the interpreter expects to come up during the conference (Kalina, 1998, p.116). In the conference interpreting setting of the European Parliament (EP), it has to be mentioned that the conference participants (i.e. mainly members of the European Parliament (MEPs)) remain fairly constant during their five-year term, and therefore interpreters in the EP can be expected to be fairly familiar with the speakers they are interpreting (although some MEPs may speak more frequently than others). The corpus material used in this thesis only includes speeches delivered by MEPs, not by external speakers. Furthermore, accredited interpreters generally work for the EP on a regular basis. Some permanent interpreters may even have more experience in the EP than newly elected MEPs, also with respect to knowledge of the terminology used in this context. Therefore, there may not necessarily be a deficit in interpreters' knowledge compared to general conference participants in this particular setting of conference interpreting. However, the *preparation strategy* is certainly a strategy also used by EP interpreters to prepare for current topics from all the European member states or international topics that may come up during the EP plenary sittings, as well as very specialised topics involving specialised terminology (e.g. new legislation). The *preparation strategy* also includes last minute speech manuscript preparation that may be available to the interpreters just before the assignment. Some of the speeches held in the EP are prepared and read-out speeches that may include more characteristics of written language, such as high lexical density and sophistication. However, it should be noted that, if these manuscripts are available to interpreters at all, last-minute preparation is not as thorough as preparation without imminent time pressure.

The interpreting strategy of *preparation* may influence other comprehension enhancing strategies, such as *inference* or *anticipation*. *Inference* is a knowledge-based strategy to "fill the gap". Logical conclusions are derived from knowledge, e.g. if

the source speaker did not explain background information explicitly. *Inferencing* also helps understanding in imperfect interpreting conditions, e.g. with background noise or a speakers' strong accent. Furthermore, *inference* can be used to overcome cultural differences and make the target text more explicit for something that was left implicit in the source (Kalina, 1998, p.116). The strategy of *inferencing* is also linked to the above-mentioned strategy of *anticipating* (Seeber, 2011), where the interpreter produces a source speech item before it was uttered by the source speaker, relying on linguistic cues such as lexical collocations, suprasegmental features, or certain syntactic structures, as well as knowledge cues (knowledge about the topic or speech context) (Li, 2015). *Anticipation* may strongly depend on the languages involved in the interpreting process (cf. result from Defrancq (2015); Donato (2003) or Bartłomiejczyk (2006)). The example in Figure 2.1 shows anticipation of the verb as shown. The verb in final position in the German source (sentence 1) is anticipated as required earlier in the English target (sentence 2):

(1) Wir glauben, dass die Delegierten ihre Entscheidung nach einer langen Diskussion treffen.

(2) We believe that the delegates make their decision after a long debate.

It can be assumed that target speech items that are predicted (or anticipated) by the interpreter will be highly expected items in context, i.e. words with low surprisal (cf. Section 3.1.1).

The last *comprehension enhancing strategy* given by Kalina (1998) is the strategy of *chunking* (also cf. Seeber, 2011). Kalina (1998) focusses on the comprehension side, where the incoming text, which is already separated into chunks by pauses and intonation used by the source speaker, is stored in the interpreter's memory in smaller units or chunks to facilitate the comprehension process of the interpreter. The interpreting product data also show signs of such chunking in pauses used by the interpreter as well as from splitting one source text sentence into several sentences in the target text. In the example, a long source sentence (sentence 3) is chunked into smaller comprehension units (sentence 4), (Jones, 1998, p.91 f.):

(3) Japan, in the light of the ruling of the international panel, and following the non-payment of the compensation by the American steel exporters, which the US authorities have not forced them to pay, despite their legal obligations and the assurances they have given, has decided to act unilaterally, which they are perfectly entitled to in the case of non-compliance with an international panel ruling - and that is the case here - by imposing punitive duties on the import of certain flat products, although long ones should remain unaffected, at least for the immediate future.

(4) The international panel has made its ruling. Compensation has not been paid by the American steel exporters. The US authorities have not obliged them to pay, although they have legal obligations and have given assurance in this respect. So Japan has decided to react unilaterally. It is quite entitled to do so, as the ruling of the international panel has not been respected. It will impose punitive duties on imports of some flat products. Long products should not be immediately affected.

Jones (1998) calls this process of splitting long and complex source speech sentences into several shorter sentences with fewer clauses "the Salami Technique". By "slicing up" a sentence, the interpreter can begin the production of a sentence without having heard the complete sentence. He gives this as advice to students of simultaneous interpreting. As a result of the Salami Technique, the interpreter's text "is slightly shorter than the original, and the way the different sentences or clauses have been related to one another has generally been simplified and streamlined" (Jones, 1998, p.93). It can be argued that chunking is not just a strategy that facilitates comprehension of the source text, but there is a clear overlap with the target text production strategy of splitting long source sentences into several shorter ones in the target. The result in the interpreting product may not always be clearly linkable to either the comprehension side or the production strategy.

There are two further strategies that can be added to Kalina's (1998) group of *comprehension enhancing strategies*. The strategies of *waiting* silently or *stalling* by repeating previously uttered material or neutral fillers (e.g. by adding a "ladies and gentlemen") in order to buy time to process more of the source text that Seeber (2011) describes in the context of syntactic asymmetry (cf. Section 2.2.1). In a small-scale German-Greek corpus study, Liontou (2011) identifies *stalling* as a general strategy that is used very frequently. It can be assumed that the addition of neutral fillers would generally consist of expected (and low surprisal) material, but, in case of repetitions, this could also lead to unexpected word sequences in context.

When aiming to detect these *comprehension enhancing strategies* in corpus data, *inference, stalling, waiting* and *anticipation* require a parallel corpus analysis in order to compare source and target texts, and *anticipation* and *waiting* in particular require time aligned data. *Chunking* can be analysed in a comparable and parallel approach. The parallel approach where source and target speech is compared, especially if time alignment is available, can be used to detect *chunking* as a comprehension enhancing strategy. A comparable perspective, comparing chunk length in original texts and interpreted texts in the same language, can be added to argue that *chunking* can also be seen as a production-oriented measure. If the chunks produced in interpreting were shorter compared to chunk length in non-mediated production in the same language, this could point to *chunking* also facilitating production in the cognitively particularly challenging setting of simultaneous interpreting.

Kalina's (1998) framework further includes strategies that aim to facilitate the

target text production. These are subdivided into *source text conditioned strategies, target text conditioned strategies, repair strategies*, and *global strategies*.

Source text conditioned strategies relate to the source text syntactic level as well as the semantic dimension. From a syntactic perspective, the interpreter can rearrange the order of the source text elements within a sentence or even on text level (*syntactic transformation*). At the micro-level, strategies such as *paraphrasing* or *sentence splitting* can be applied (Kalina, 1998, p.118). In a qualitative study, Goldman-Eisler (1971) observes that interpreters tend to ignore input chunks, and impose their own segmentation on their output. Interpreters even mostly start to produce their output before an input chunk has fully been uttered, as observed in 90% of the material they studied. In their data-driven comparative study, He et al. (2016) clearly find traces of *sentence splitting* in simultaneous interpreting, indicated by more connectives and more sentences in a text chunk. Furthermore, they find repeated words and frequent use of demonstratives resulting from transforming clauses into independent sentences. He et al. (2016) even show great agreement among the interpreters on the chosen segmentation (73.7% agreement for two to four interpretations for the same speech). *Sentence splitting* as a micro strategy of *syntactic transformation* can be seen to reduce cognitive load as source text elements need to be stored in memory for a shorter period of time. As mentioned above and although *sentence splitting* is clearly seen as a target text production strategy, there is an overlap with *chunking* as comprehension enhancing strategy and the results in the product data are not always clearly distinguishable as source or target text conditioned strategy. Regardless of whether it results from comprehension or production processes, the resulting reduced syntactic complexity can clearly be linked to simpler target output.

According to Kalina (1998), *syntactic transformation* can be applied to buy time and find a solution to a difficulty with a source text component, to reduce the syntactic complexity of the source text, or even to adhere more to the conventions of the target language. However, it requires sufficient available cognitive capacity, because maintaining source text elements in memory and reusing these elements at a later stage requires additional cognitive effort (Kalina, 1998). In fact, Liontou (2011) finds that the strategy of rearranging source text items is only used infrequently in her qualitative corpus study on German-Greek interpreting data. The languages involved in the interpreting process may play a role here. When interpreting from head-final languages (e.g. Japanese or German) into head-initial language (for example, English or French), interpreters may use *passivisation* as a means to change the syntactic structure. This was shown by He et al. (2016) for Japanese to English interpreting, who see passivisation as an acquired skill that can speed up the interpreting process for the interpreting direction Japanese to English.

According to Kalina (1998), the cognitively more beneficial syntactic strategy is *transcoding*. It is an almost word-for-word interpretation of the source text that can be applied if the target language norms allow for this strategy to sound natural in the target text (Kalina, 1998, p.118). Kalina specifically refers to lists of names or

numbers, where this strategy can be applied very successfully. Moreover, *transcoding* can also be applied when the interpreter aims to avoid the additional cognitive load that would be necessary to apply *syntactic transformation* (Kalina, 1998, p.119). However, this can lead to patterns in the target speech that - although possibly still acceptable in the target language - may be used in different frequencies than in original production. Such syntactic shining-through of the source language can lead to target output that is less expected by the target audience and could potentially also increase syntactic complexity.

On the semantic level, semantic *compression* can be used as a further *source text conditioned text production strategy*. Although *compression* may not always be used as a strategy, as it can also be the result of a source text component not being heard or understood by the interpreter, intended and therefore strategic *selection* and *deletion* refers to selecting main elements of the source text and omitting redundant or less important components to still achieve the communicative goal (Kalina, 1998, p.119ff.). Other scholars refer to such an incomplete rendition of source information as *omission* (Gile, 1995; Setton, 1999; Korpal, 2012; Jones, 1998) or *skipping* (Al-Khanji et al., 2000). *Omission* is sometimes discussed in the context of errors or interpreters facing cognitively difficult conditions (Gile, 1995, 1999; Setton, 1999). However, it is also seen from a pragmatic perspective with *omission* being treated as a conscious strategy rather than a mistake (Pym, 2009; Korpal, 2012). In fact, Korpal (2012) shows that professional interpreters do not statistically omit more in cognitively challenging situations, such as source speeches with fast delivery rates, but rather tend to omit less informative information in faster interpreting settings. This, along with results from a questionnaire about reasons for omitting, points to a conscious choice of leaving source speech content out in the interpreter's output. Al-Khanji et al. (2000) describes *skipping* as a compensatory strategy, where the interpreter avoids single words due to either (a) incomprehensible input, i.e., the interpreter does not know the meaning of a term, (b) the interpreter consciously decides that a term is repetitive and the overall meaning is covered, or (c) the interpreter is lagging behind and decides to leave out content to reduce the lag (Al-Khanji et al., 2000, p.553). Al-Khanji et al. (2000) find *skipping* the most prominent strategy used by interpreters in a qualitative study with interpretation from English to Arabic. Nevertheless, some occurrences of *omissions* also result from a failure in text comprehension. This can be observed when sentences are not completed and larger units of text are omitted. This type of semantic compression was found to be the fourth most used category in the study led by Al-Khanji et al. (2000), referred to as *comprehension omission*.

When assessing the effects of conscious or unconscious *omissions* in the interpreting product, it can be assumed that leaving out some source text content can have a simplification effect or exactly the opposite. Omitting redundant content or less important utterances (such as omitting filler items) can render the target text more informationally dense and, therefore, less simplified. The same is true if the interpreter

leaves out connective items leaving logical relations more implicit. In this case, the interpreter's audience is required to infer the logical relations themselves, requiring more effort on the audience side. If the interpreter omits information such as giving only 2 out of 5 items in a list, skipping asides, or digressions, the target speech may become less informationally dense, and therefore more simplified compared to the original utterance. However, this can only be said if the interpreter is still able to produce coherent target speech.

Jones (1998) links *omissions* to fast delivered source speeches as well as prepared and read-out speeches. It should be considered that the corpus data used in this thesis, European Parliament plenary speeches, are generally delivered faster than in other conference interpreting settings and some of the speeches are prepared and read-out (Monti et al., 2005).

Semantic compression can also be achieved by using a term that is more general, i.e. higher in the hierarchical structure (Kalina, 1998, p.120). Such *generalisation* is used as an approximation strategy in which the closest possible solution is selected instead of a direct equivalent and the target text becomes deliberately more vague (Kohn and Kalina, 1996). Al-Khanji et al. (2000) refer to this strategy as *approximation* of individual source speech expression and find it the second most frequently used strategy in their study. Al-Khanji et al. (2000) further refer to *filtering* or *summarising* (Li, 2015) as means of semantic compression where the length of an utterance is compressed while preserving the semantic content of a message. This strategy is used as the third most frequently used strategy (Al-Khanji et al., 2000).

In the example sentences 5 and 6 (examples from (Jones, 1998, p.101)), a list of several specific items (sentence 5) is replaced with a term higher up in the semantic hierarchy (sentence 6). Such *generalisation* and *summarising* can be assumed to free cognitive resources by having to retrieve one instead of several lexical items, as well as by using words that are higher up in the semantic hierarchy. Furthermore, this strategy can be used to save time if the interpreter is lagging too far behind the speaker.

(5) People take it for granted now to have a fridge and a freezer, the dishwasher and the washing machine with a spin dryer, a cooker and a vacuum cleaner.

(6) People take it for granted now to have all household electrical appliances.

Camayd-Freixas (2011) and Al-Khanji et al. (2000) find that interpreters tend to deliver the gist of a sentence rather than to give the nuanced meaning of each word. This kind of semantic *compression* results in the choice of more frequent words because their retrieval time is faster (Dell and O'Seaghdha, 1992; Cuetos et al., 2006). He et al. (2016) find evidence for this strategy in their corpus study with fewer stem types, token types, and content words in interpreting which they link to *generalisation*. Furthermore, they see this as a by-product of humans' limited working memory.

Kalina (1998) further lists *simplification* as a separate strategy and lists simplifying syntax (e.g. sentence splitting, paraphrasing or restructuring), lexis, and style. Kalina (1998) does not specify in more detail what simplifying includes in particular, and a clear overlap with other strategies mentioned above can be seen: for example, simplifying syntax may already be covered by *syntactic transformation*, or some lexical aspects by other semantic compression strategies such as choosing more general high-frequent words. In interpreter training, the advice is given that, under some circumstances, it may be advisable or even necessary to simplify (Jones, 1998). This refers to highly technical material, where it can happen that, despite best preparation efforts, the interpreter cannot provide all technical specifications given in the source, or the interpreter may even be advised to choose to simplify as a means of audience design to maximise communication (Jones, 1998, p.98). This clearly overlaps with some aspects of semantic *compression*, e.g. *generalising* or *approximating* as well as *filtering* or summarising and compressing, however, not maintaining the semantic content, but instead reducing the information transmitted.

As the definition of *simplification* is not clear cut, it is difficult to trace it fully in the interpreting product. It is often traced via reduced lexical variety in the target text (Baker, 1993, 1996; Laviosa, 1998; Kajzer-Wietrzny, 2012). In fact, a trend of more repetitive and less diverse lexis, a narrower range of vocabulary, and lower sophistication in simultaneous interpreting than original spoken language can generally be observed (Russo et al., 2012; Kajzer-Wietrzny and Ivaska, 2020; Xu and Li, 2022). However, there are some exceptions to this trend (Ferraresi et al., 2018; Dayter, 2018, among others). Lexical density as a further simplification measure shows no general trend of simplification in interpreting, but rather densification (Kajzer-Wietrzny, 2015; Sandrelli and Bendazzoli, 2005; Ferraresi et al., 2018; Dayter, 2018; Lv and Liang, 2019) (cf. Section 2.4.2 for a detailed overview). Such results show interaction with other semantic *compression* strategies such as avoidance of repetition and redundancies.

Kalina (1998) states that these semantic *compression* strategies are *emergency strategies* that can be applied when the flow of information that the interpreter has to process exceeds the processing capacities of the interpreter. This may be due to fast delivery, complex structure of source or informationally dense input, but also due to fatigue, background noise, or other source text independent factors. To study whether these *source text conditioned production strategies* are really an emergency operation or rather a means of coping with the generally cognitively challenging conditions in simultaneous interpreting, a parallel corpus study is required to see potential triggers within the source text, combined with a comparable corpus study to see if interpreters overuse certain phenomena more than original speakers in comparable settings.

In addition to these *source text conditioned production strategies* within the *target text production strategies*, Kalina (1998) lists *target text conditioned strategies* that are

conditioned by the source text to some extent. However, the target text production is more in the focus. Such strategies include the decision to shorten or lengthen *EVS* or *presentation strategies* that refer to extralinguistic presentation means such as intonational choices, pause patterns, or adaptation of speech production rate to source discourse delivery rate and/or listener's requirements. Furthermore, the interpreter can also choose to *create coherence in the target text* by inserting additional connective elements, or to *compress* or *expand* the target text by deleting or adding redundancy or adding textual material for clarification purposes (audience design) (Kalina, 1998, p.119). Kalina (1998) lastly adds *stylistic decisions* to the *target text conditioned strategies* such as selecting the best expressive means, independently of the source text (Kalina, 1998, p.119).

To complete the set of strategies available to the interpreter, Kalina (1998) adds *repair strategies* as the conscious decision to correct or not correct a production error, as well as the *global strategy* of *monitoring* where the interpreter checks that their target output is coherent (Kalina, 1998, p.120).

To summarise the overview of strategies that may be used by interpreters to cope with the cognitively challenging task of simultaneous interpreting, it is important to recall that the interpreter's business is not words, but ideas or message elements. "The interpreter is continually involved in evaluating, filtering, and editing (information, not words) in order to make sense of the incoming message and to ensure that his output, too, makes sense" (Henderson, 1982 cited in Al-Khanji et al., 2000). The more automatised the strategies become on both the comprehension and production side, the more they are assumed to reduce the cognitive load during the interpreting process (Kalina, 1998, p.121). Li (2015) also observes that when strategies "become unconscious or automatic ... they reduce the cognitive load because only if strategies are activated automatically will the interpreter overcome his or her capacity limitations and make best use of available processing capacity" (Li, 2015, p.172). It is important to stress that no strategy fits all the time, but instead the strategies interact and interdepend. Depending on the source input, interpreters may opt for one strategy or another, or a combination of several strategies.

The overview of strategies available to interpreters given in this section shows what means of coping with cognitive constraints in interpreting can be used in simultaneous interpreting. These strategies are also taught to interpreting students (Jones, 1998) and some of them have been observed in the interpreting product. It is debatable whether some strategies, such as semantic *compression*, are only used when the interpreter risks exceeding processing capacity (Kalina, 1998; Gile, 1995), or whether they are rather a method that is generally applied to deal with the cognitively challenging process of simultaneous interpreting. In order to link the effects in the product data to results of cognitive load in interpreting, parallel as well as comparable corpus analyses are required. Different approaches to trace the effects of cognitive load in simultaneous interpreting via product data are given in the next

section (Section 2.2.3). Furthermore, Section 2.4 of this thesis will show what is empirically known about interpretese so far.

2.2.3 Approaches to measure cognitive load in simultaneous interpreting

Besides theoretical interpreting models acknowledging the importance of high cognitive demands (Section 2.2.1) and available strategies to deal with limited cognitive resources (Section 2.2.2), interpreting scholars also approach cognitive load or overload via the interpreting process and product data. Process data include *pupillometry* data (Seeber, 2011) and interpreters reporting on the individually perceived load during interpreting via *retrospective protocols* (Ivanova, 2000; Bartłomiejczyk, 2007; Gumul, 2020, amongst others). In approaches to measure coginitive load in interpreting product data, the occurrence of disfluencies in corpus data is generally used as a link to cognitive load (Plevoets and Defrancq, 2016, 2018b; Defrancq and Plevoets, 2018; Plevoets and Defrancq, 2020; Chmiel et al., 2023; Jiang and Jiang, 2020, amongst others). Cognitive load in this context can be defined as "the proportion of an interpreter's limited cognitive capacity devoted to performing an interpreting task in a certain environment" (Chen, 2017, p.643).

Approaches to cognitive load via interpreting process data

An approach to measure cognitive load by tapping into the interpreting process is *pupillometry* (Seeber and Kerzel, 2012a). In a controlled experiment using task-evoked pupillary responses (TEPRs), Seeber and Kerzel (2012a) measure *pupil dilation* as an indicator of cognitive activity. In their setup, German source input has to be interpreted into English, with symmetrical structures used as the baseline setting compared to an asymmetric setting where a verb-final structure has to be transformed into a verb-initial structure. Their results show higher cognitive load generated by the asymmetric sentence condition - in line with the CLM proposed by Seeber (2011). Although the experiment does not allow for the resulting increase in cognitive load to be linked to a specific interpreting strategy (waiting, stalling, chunking, or anticipating), the results from this experiment also give other interesting insights. They show differences in cognitive load for interpreting sentences without context compared to sentences given with context, with higher cognitive load for sentences given without context compared to context-embedded sentences. Their results show that cognitive load in particular difficult conditions, as theoretically presumed, can be measured during the interpreting process. Seeber (2011) as well as studies by Tommola and Niemi (1986) and Hyönä et al. (1995) show the applicability of pupillometry for simultaneous interpreting process research within a strictly controlled experimental setting. Unfortunately, the method does not seem to be applicable for larger units of speech with no measurable difference in mean pupil dilation for texts of 340-words length

(Schultheis and Jameson, 2004 cited in Seeber, 2013) and therefore cannot be applied to larger units of text, including more than one controlled presumed interpreting difficulty. Moreover, it cannot be linked directly to actual interpreting product data from real live settings and reflects the various types of simplification that may be found in the interpreting product.

A second approach to tap into cognitive processes during simultaneous interpreting is asking interpreters about their output, as well as the problems, solutions, and individual decisions made at a certain point during the interpreting process. *Retrospection* can be used to gain insight about conscious and controlled processes (Ivanova, 2000). When using *retrospection* in simultaneous interpreting, subjects are asked questions after they completed the interpreting task to avoid interrupting the interpreting process as such. Due to the fairly long time interval between the actual interpreting task and the retrospection task, different types of cue can be given to the interpreter to improve recall (Ivanova, 2000): (a) transcript of the source text or (b) transcript of the source text and recording of ones own interpreting product. It can also be helpful to remind the interpreter of their behaviour during the interpreting process, such as exited gesticulation, facial expression, hesitations, or inaccuracies in their performance (Ivanova, 2000, p.33).

When assessing the *retrospective protocols*, it has to be considered that the obtained vocalisations are based on information on the thought process that is still stored in memory and therefore subject to forgetting or inferring information instead of retrieving it from memory. This may influence the accuracy and completeness of the data obtained. Instructions given to the interpreter within the retrospection task can also lead to bias, as interpreters might include speculation in their report. Therefore, the accuracy of the information obtained via *retrospection* may be reduced (Ivanova, 2000). Furthermore, the type of cues given to the interpreter, personal determinants, or even the relationship between the interpreter and the researcher may influence the data obtained (Ivanova, 2000). If participants are confronted with their own output, they may try to comment or justify their interpreting choices (Dimitrova and Tiselius, 2009). Gumul (2020) shows that the methodological issue of type of retrieval cues offered may not be so critical, as idiosyncratic differences have a greater impact on retrospective reports. All these factors make some researchers argue that retrospection does not provide the objectivity required to assess cognitive load in a reliable way (Seeber, 2013). Others conclude that if researchers are aware of these factors, "verbal reports elicited with care and interpreted with full understanding of the circumstances under which they were obtained, are a valuable and thoroughly reliable source of information about cognitive process" (Ericsson and Simon, 1980, p.247).

Ivanova (2000) shows that professional interpreters report primarily on problems and monitoring observations related to high-level text processing and translation. Problems are reported resulting from comprehension difficulties such as constructing

a coherent representation of the source language chunk, but also related to the translation process as such where they referred to problems in selecting the appropriate target language equivalent for the target language context for chunks of the source message. This particularly relates to comprehending proper nouns even though the names were presented to the subjects before the interpreting task.

In a study with interpreting students, Bartłomiejczyk (2006) shows that participants can recall strategic choices they made at a certain point during interpretation. However, since some strategies may be automated, the method of retrospection should be combined with a product analysis to compare instances of reported strategy and actual strategy application in the interpreting output. This approach is taken by Gumul (2019), who aims to tap into cognitive processes by combining a process analysis via *retrospective protocols* with a product analysis of manual corpus comparison of source and target texts.

Gumul (2019) takes the following indicators of problems in product data and links them to cognitive effort: anomalous pauses, omissions, repairs, grammatical errors, mispronunciations, hesitation markers (filled pauses) and false starts. Gumul's retrospective protocols show that, in order to avoid an excessive load on short-term memory, interpreters shift information to a cognitively more favourable position: a list of items might be produced early, with very short EVS, and information that initiated a list in the original speech might be added at the end of the list in the interpreted output. Focusing on three types of disfluencies as indicators of increased cognitive load, hesitation markers, false starts, and anomalous pauses that exceed two seconds, Gumul (2017) also shows that processing problems can lead to the addition of elements: repetition of elements, addition of modifiers and quantifiers, and specification of meaning. Although this type of explicitation was found to be primarily a process-oriented strategy of interpreters coping with input that creates high cognitive load, rather than audience design (Gumul, 2017, p.282ff), the result in the interpreting product can still be seen as simplification, as more repetitive and explicit target text would be simpler for the audience to process. Participants in Gumul's study were interpreting students and such explicitation patterns are found to be expertise-related with professional interpreters using less of this strategy than interpreting students (Tang, 2018). Nevertheless, the retrospection method can link the result in the product data to triggers of experienced processing difficulties. However, Gumul (2019) also shows that the cognitive load experienced as reported by the interpreters in the retrospection and traces of cognitive load in the interpreting product do not always coincide. Although interpreters may experience and report increased cognitive load during the interpreting process, no problem indicator may be observable in the interpreting product.

Retrospection can be applied to shed light on the strategies used by interpreters when faced with processing difficulties. However, it can only trace the conscious decision. Unconscious and automated decisions cannot be reported. Furthermore,

not all instances of experienced cognitive load during simultaneous interpreting lead to direct traces in the interpreting product. This thesis explores whether interpreters use simplification in order to cope with cognitive constraints during the interpreting process. This may result from automatic, and sometimes unconscious processes. The experimental setup only allows for investigation of a very limited amount of data (regarding subjects as well as length of speech investigated). Corpus-based approaches overcome both of these issues.

Corpus-based approaches to cognitive load in interpreting

Various studies aim to capture cognitive load in interpreting using corpus-based approaches, mainly using disfluencies in the interpreting product as indicators of cognitive load linked to potential triggers of high load in source and/or target such as triggers of lexical problems, syntactic complexity or speed of delivery (Plevoets and Defrancq, 2016, 2018b; Defrancq and Plevoets, 2018; Plevoets and Defrancq, 2020), word frequency and cognate status (Chmiel et al., 2023) or dependency distance measures (Jiang and Jiang, 2020). Other approaches do not use indicators of local cognitive load, but rather use measures related to phenomena that are distributed throughout the interpreting product, such as syntactic complexity (Liang et al., 2017) or lexical simplification (Lv and Liang, 2019). They are discussed in Section 2.3.2 and 2.4.2, respectively.

Speech planning constraints are assumed to manifest themselves in fluency disruptions (Pöchhacker, 2004). Fluency disruption of speech and filled pauses in particular are associated with processing difficulties and are therefore investigated as a direct link to cognitive load (Defrancq, 2018; Kajzer-Wietrzny et al., 2024). They are linked to difficulties in comprehension, access to translation equivalents, or production challenges (Cecot, 2001). Filled pauses are assumed to be linked to processing difficulties and tend to occur particularly in the context of new information (Clark and Fox Tree, 2002) or heavy constituents (Arnold et al., 2000). The occurrence of filler particles may also be linked to expectedness / unexpectedness of a lexical word in its context (i.e. its surprisal - see Section 2.3.1) as well as with syntactic integration and storage costs (see Section 2.3.2) (Dammalapati et al., 2021). More disfluencies are observed at the beginning of long and complex sentences (Clark and Fox Tree, 2002; Konopka, 2012)). Silent pauses can also signal comprehension problems or difficulties in retrieving target language equivalents (Chmiel et al., 2017). Generally, fewer silent pauses are observed in interpreting compared to monolingual speech, but these fewer silent pauses tend to be longer in interpreting (Tissi, 2000). In general, in interpreting, filled pauses tend to be more frequent than silent pauses with more frequent filled pauses in higher source text delivery rate contexts (Cecot, 2001). These differences in the distribution of filled and silent pauses in interpreting compared to monolingual speech point to differences in speech planning activities. On the one hand, pauses

may be related to a high cognitive load, but can also be used strategically to gain time (Chmiel et al., 2017).

Several corpus-based interpreting studies investigating cognitive load in interpreting therefore use silent and/or filled pauses to investigate triggers of these fluency disruptions that cause high cognitive load. Plevoets and Defrancq investigate the triggers of disfluencies in several studies (Plevoets and Defrancq, 2016, 2018b; Defrancq and Plevoets, 2018; Defrancq, 2018; Plevoets and Defrancq, 2020). The authors analyse a number of known load factors in interpreting input and output, as well as their association with filled pauses. They use delivery rate, frequency of numbers, lexical density, formulaicity, average sentence length (only in their earlier study, Plevoets and Defrancq (2016)), and number of subordinations (in Plevoets and Defrancq (2018b); Defrancq and Plevoets (2018); Plevoets and Defrancq (2020)) as indicators that are known to affect cognitive load during comprehension and are likely to also affect load during production. The authors assume that "producing complex utterances is also more challenging than producing simple ones" (Plevoets and Defrancq, 2020, p.21). Plevoets and Defrancq (2020) argue that cognitive load during interpreting is not only affected by the characteristics of the input speech, but the interpreting strategies used. Interpreting quality or linguistic features of the target text can also influence the cognitive load during interpreting. A more simplified target speech would also reduce cognitive load indicated by fewer filled pauses in such contexts.

Gile (2008) proposes the idea that cognitive load in interpreting may be imported across sentence boundaries into the following sentence. This idea of "imported load" or "knock-on effects" (Gile, 2008) is also referred to as a "spillover effect" or "exported load" (Chmiel et al., 2023) in the literature. It is also described as "a considerable increase in the cognitive load downstream" (Seeber and Kerzel, 2012b, p.229). However, the existence of imported load is largely based on anecdotal evidence and has not been empirically proven (Plevoets and Defrancq, 2020; Chmiel et al., 2023).

The idea of "imported load" sounds reasonable in the context of the interpreting process and potential load distribution. Interpreters generally lag behind the source speaker. This time distance between a source speech unit and the translation equivalent in the target speech (EVS) is generally between 1 and 3 seconds (Defrancq, 2015). Usually, the source sentence is completed before the target sentence. This is illustrated in Figure 2.2, showing that the comprehension phase of a sentence (sentence 2) usually overlaps with the production phase of the previous sentence (sentence 1) (Plevoets and Defrancq, 2020, p.22). Plevoets and Defrancq (2020) specifically investigate whether the delay in the production load of the previous sentence has an impact on cognitive resources management that goes beyond the internal production load of the sentence.

Sentence processing research gives evidence that comprehension load is not evenly distributed across a sentence, but instead peaks at the beginning and end of a sentence

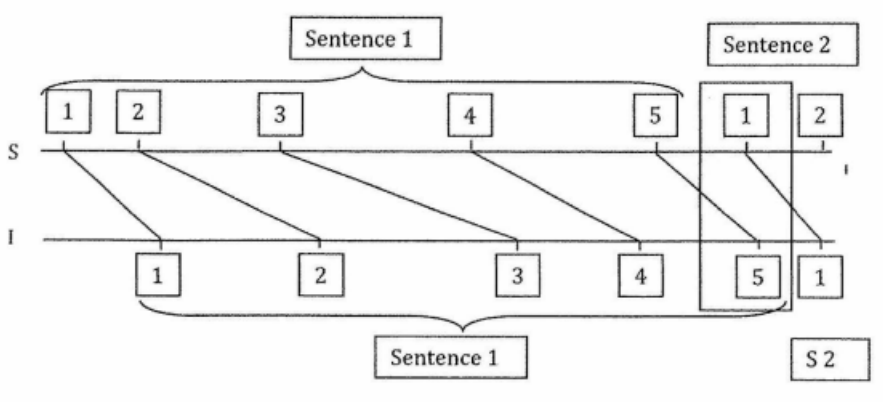


Figure 2.2: Representation of an interpreting performance with EVS and imported load, S - speaker, I - interpreter (Plevoets and Defrancq, 2020, p.22).

(Granier et al., 2000) as indicated by arrows in the top row in Figure 2.3 (Plevoets and Defrancq, 2020, p.23). The cognitive load resulting from sentence production generally peaks before and at the beginning of a sentence, as evidenced by a high number of disfluencies at the beginning of a sentence (Barr, 2001). These assumed peaks of production load in sentence production are indicated as arrows in the lower row in Figure 2.3 (Plevoets and Defrancq, 2020). Although the representation does not consider an additional load from interference between activities that are conducted at the same time, Figure 2.3 shows the potential overlap of load peaks at the sentence boundary.

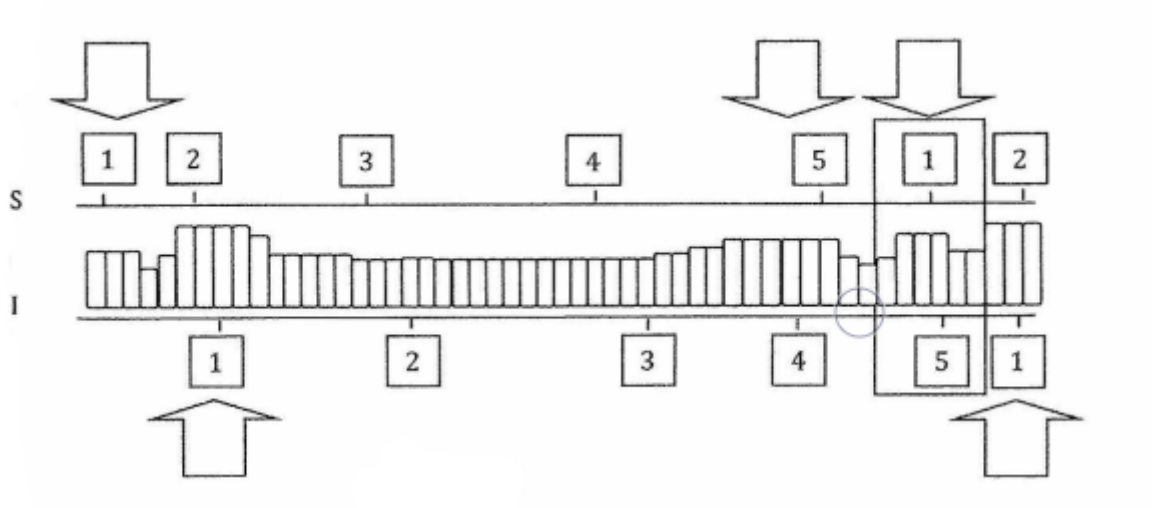


Figure 2.3: Representation of interpreters' theoretical load peaks and load evolution (Plevoets and Defrancq, 2020, p.23).

In a corpus-based approach, Plevoets and Defrancq (2020) use statistical analysis via a generalized additive mixed-effects model (GAMM) to predict the occurrence of filled pauses (response variable) in the interpreting product. As predictor variables, they use lexical properties of cognitive load (lexical density, frequency of numbers,

formulaicity), syntactic complexity (percentage of subordinations), and delivery rate in interpreted source and target texts. Furthermore, they also compare their results for the Dutch interpreting data to a comparable monolingual Dutch dataset. To investigate the potential load being imported into the following sentences, the frequency of disfluencies is linked not only to observed predictor variables in the current sentence but also to the occurrences in the three previous sentences. The study shows that only filled pauses in the current sentence can be predicted by four of their ten predictor variables for cognitive load: lexical density of source, formulaicity of source and target, as well as the number of subordinates in the source. However, filled pauses in a sentence cannot be predicted by the imported load of previous sentences (Plevoets and Defrancq, 2020). The results for the comparable monolingual dataset are reversed. For non-mediated speech, the authors indeed find a knock-on effect with filled pauses being predictable using potential triggers of cognitive load from previous sentences. As no imported load in interpreting is observed, but only in the non-mediated context, the authors reason that unlike speakers, interpreters "are able to reset their working memories at the sentence boundary" of their own production (Plevoets and Defrancq, 2020, p.35). Interpreters are able to "free their working memory at specific points in the text to be able to cope with subsequent input" (Plevoets and Defrancq, 2020, p.39). This finding is crucial for the present study as it relates to an approach investigated in this thesis. Chapter 6 investigates the length of dependency relations in interpreting, that is, the distance between an element in the sentence and its head. At the end of a sentence, all syntactic relations are closed; the speaker no longer needs to keep open relations in memory to be resolved. Moreover, the results of Plevoets and Defrancq (2020) show that sentences can be used as units to investigate the effects of cognitive load in interpreting as cognitive load is unlikely to be transferred across sentence boundaries.

The detailed results of Plevoets and Defrancq (2020) and their previous studies are also very informative as to what triggers filled pauses as indicators of high cognitive load. High lexical density of the source text is associated with filled pauses in interpreting (Plevoets and Defrancq, 2020). This confirms previous studies showing that filled pauses are caused by problems with lexical retrieval, e.g. the authors showed that filled pauses in interpreting are more frequent with high lexical density in the interpreting product (Plevoets and Defrancq, 2016) or the source text (Plevoets and Defrancq, 2018b) is higher. This is in line with results of a further study by the same authors which is on filled pauses within compounds as highly dense structures. They show that retrieval of compound lexemes is particularly cognitively demanding with significantly more intra-compound filled pauses in interpreting compared to monolingual speech (Defrancq and Plevoets, 2018). Taking into account that interpreters tend to produce more filled pauses overall, the authors show that compound retrieval generates cognitive load in interpreting, including frequent compounds. Cross-linguistic differences in the order of components parts of the compounds, and therefore an additional effort to restructure the components, also seem to have an effect (Defrancq

and Plevoets, 2018).

On the other hand, when the interpreter faces highly expected input, such as highly formulaic source speech, fewer filled pauses are observed (Plevoets and Defrancq, 2020, p.30). This is an expected result on the comprehension side as formulaic sequences are known to ease processing, e.g. shown in reading studies with shorter fixations (Conklin and Schmitt, 2008). They are assumed to be processed as whole sequences rather than word-by-word and therefore require less working memory capacity (Tremblay et al., 2011; Tremblay and Baayen, 2010). While source text formulaicity shows a moderate negative relationship with filled pauses in the target, high target text formulaicity is also linked to fewer filled pauses in the target text (Plevoets and Defrancq, 2020, 2018b). The processing advantage of formulaic sequences can therefore be observed not only for comprehension but also for production processes.

The authors link their results for lexical density and formulaicity to interpreters being able to "balance comprehension load by reducing production load" (Plevoets and Defrancq, 2020, p.30). When more input load is perceived, the output load is reduced by producing more standard material. This is an interesting observation, as it can be hypothesised that this holds true not only for lexical density and formulaicity but potentially also to other mechanisms that can be used to simplify the target output.

Furthermore, results on syntactic subordination are particularly interesting in the study by Plevoets and Defrancq (2020). In monolingual speech, syntactic subordination is a known predictor of processing costs (Gordon and Luper, 1989; Osborne, 2011). However, in interpreting research, Dillinger (1994) and Setton (1999) do not observe strong effects of syntactic embeddedness of the source on interpreters' output. Regarding subordination in their study, Plevoets and Defrancq (2020) only find subordination in the source as a predictor of filled pauses in the target texts. However, the results for this predictor are less clear than for other predictor variables. Higher syntactic complexity in the source does not necessarily increase the predictability of filled-pause load. Subordination in target (interpreters production) and complexity in previous source sentences is not a predictor of increased cognitive load indicated by filled pauses in the target (Plevoets and Defrancq, 2020, p.31).

An unexpected result is also found for the frequency of numbers and their effect on disfluencies (Plevoets and Defrancq, 2020). Numbers are often referred to in the literature as difficult to be interpreted and increasing cognitive load (Gile, 2009). They are assumed to be low in predictability, low in redundancy, and carrying high informational content (Mazza, 2001; Pinochi, 2009). However, different categories of numbers may be processed differently. Numbers can have a pure arithmetic value (e.g. 5 or 218), express an order of magnitude or approximation (e.g. millions), be combined with a unit of measurement (e.g. metres, kilogrammes) or other referents (e.g. measure of time), or can be compared to other numbers with the same or different properties (Jones, 1998). Kajzer-Wietrzny et al. (2024) show that differences in the types of numbers affect the occurrence of disfluencies with ranges of numbers

being less associated with disfluencies, whereas numbers as part of names, decimals, or numbers with more than four digits are very likely to cause a fluency disruption. Furthermore, if more than one number occurs in a sentence, interpreters are also likely to produce a disfluency (Kajzer-Wietrzny et al., 2024, p.13). Even if numbers are omitted, fluency disruptions in interpreters' output are observed (Kajzer-Wietrzny et al., 2024).

However, Plevoets and Defrancq (2020) do not find a statistical effect for filled pauses related to numbers. A potential reason for this may be that they do not differentiate between the different types of numbers that may influence the likelihood of a fluency disruption. The authors explain their findings by hypothesising that this may be due to omissions of numbers in the target or interpreters having access to supporting documents during the interpreting process (Plevoets and Defrancq, 2020).

To conclude the study by Plevoets and Defrancq (2020), no traces of imported load or a spillover effect are observed across sentence boundaries. With continuous new input, interpreters have to be able to free their cognitive resources as quickly as possible to be able to process new input. The study does shed light on individual load factors in simultaneous interpreting. However, some potentially influencing factors such as errors, omissions, and infidelities in the interpreting product were not included in the statistical model, and therefore the authors cannot completely rule out the existence of a spillover effect (Plevoets and Defrancq, 2020).

Chmiel et al. (2023) also set out to investigate a spillover effect, resulting in a decreased interpreting performance downstream. They use a different corpus-based approach, investigating cognitive load via fluency disruption caused by source text word frequency and cognate statues. Based on European Parliament interpreting data for Polish and English (PINC, (Chmiel et al., 2022)), additional variables, such as interpreting direction (interpreters working B to A – working language interpreted into native language) and A to B (source speech in the native language interpreted into working language), are considered in the study. Low frequency words are assumed to trigger high cognitive load (Chmiel et al., 2023). This is operationalised as disruption of fluency measured via length of silent and filled pauses within a time frame of 10 seconds after the occurrence of a low frequency word in the source text. Therefore, current cognitive load (around 4 seconds following the occurrence of a critical word in the source) as well as imported load from previous segments to the current occurrence (up to 4 seconds following the occurrence of a critical word in the source) and exported to later stages (more than 4 seconds following the occurrence of a critical word in the source) are compared.

The study selects lexical words with assumed differences in cognitive processing. High-frequency cognates, i.e. words that share form and meaning across the languages, are assumed to be easily retrievable in the target language (Chmiel et al., 2023). These high-frequency cognates are compared to high-frequency non-cognates (language-unique lexical words) and low-frequency non-cognates. Generally, the au-

thors assume that low-frequency words require more processing capacity as their semantic features are more difficult to access compared to high-frequency words. Furthermore, the search for translation equivalents is assumed to be more difficult for low frequency words (Chmiel et al., 2023).

Furthermore, the study considers two important factors in cognitive processing: the frequency effect and the cognate facilitation effect (Chmiel et al., 2023). The general word frequency is considered to be one of the most important factors in lexical processing times, with high-frequency words processed faster and more accurately than low-frequency words (Brysbaert et al., 2018; Weber and Crocker, 2012). Chmiel et al. (2023) rely on a robust frequency effect that is observed in different translation tasks, including interpreting (Christoffels et al., 2006), sight translation (Su and Li, 2019), and reading times for translations (Ruiz et al., 2008). The cognate facilitation effect can be observed through faster processing of words that share form and meaning across languages in cross-lingual speech production tasks (Broersma et al., 2016; Kroll et al., 2002; van Hell and de Groot, 2008). This can be observed for interpreters and non-interpreting bilinguals in both translation directions (from and into the native language) (Christoffels et al., 2006; Lijewska and Chmiel, 2014). In line with Ploevoets and Defrancq's (2020) findings, Chmiel et al. (2023) show that lexical frequency triggers current load in the interpreting product, but they do not find any evidence of load exported downstream in the interpreting process. Their detailed results on frequency and cognate status are particularly informative. An increase in current cognitive load for words with low frequency is found, but not for high-frequency words, including both cognate and non-cognate words. Therefore, word frequency can be used as a predictor of the current cognitive load manifested in disfluencies within the time interval directly following the interpreter having heard the word in the source, indicting current load (Chmiel et al., 2023, p.12). Interestingly, a cognate facilitation effect was found only for high-frequency cognates compared to non-cognates in the interpreting direction B to A (working language into native language) (Chmiel et al., 2023, p.13). A directionality effect was also found with overall more disfluencies in B to A interpreting than A to B interpreting, indicating higher load for production in the native language and comprehension in the non-native language than in the opposite direction. Particularly, low-frequency words in B to A interpreting triggered current cognitive load, whereas interpreting cognates resulted in lower current cognitive load in A to B interpreting. The authors link their observations to comprehension difficulties in the non-native language, suggesting that processing of non-native items is more demanding than processing words in the native language (Chmiel et al., 2023, p.14). Such an effect of words being easier to be activated in the non-native language in response to the native language than in the reverse direction has also been found in translation priming studies (Finkbeiner et al. 2004, Schoonbaert et al. 2011, cited in Chmiel et al., 2023).

The authors did not find any evidence of exported load to later stages in the interpreting process triggered by lexical frequency (Chmiel et al., 2023). However, they

do not rule out the existence of such a spillover effect, which might be observable via other auditory measures, such as stress measured via fundamental frequency (Chmiel et al., 2023, 2024).

Both studies Plevoets and Defrancq (2020) and Chmiel et al. (2023) use disfluencies to investigate current and imported load. However, several studies show that disfluencies occur before an upcoming difficulty (Dammalapati et al., 2021). Pollklläsener et al. (2025) also show that filled pauses tend to occur before an unpredictable lexical item and therefore are an indication of high cognitive load before, rather than after the occurrence of a word with high retrieval costs. Therefore, it is questionable whether, if a spillover effect exits, disfluencies linked to a specific time interval are the appropriate signal in the data to investigate cognitive load being imported to the downstream interpreting process.

To conclude, the studies that investigate evidence of imported load during the interpreting process, we quote Shao and Chai (2020). In contrast to the studies mentioned above, the authors do not use fluency disruptions as indicators of increased cognitive load, but use an experimental setup measuring imported load via number of chunks or the amount of information held in working memory. Chunks of a previous sentence that are not yet produced by the interpreter when a new input chunk or sentence is heard are defined as imported load. The results obtained in this experimental setup are in line with the corpus-based approaches mentioned above, with no evidence for a spillover effect found (Shao and Chai, 2020). The authors link this finding to the coping tactics acquired by professional interpreters due to training and practice (Shao and Chai, 2020).

As a last approach using disfluencies to tap into cognitive load, we cite a study on syntactic complexity. (Shao and Chai, 2020) use an experimental approach to link syntactic complexity, operationalised via dependency distances, i.e. the linear distance between two syntactically related words and their maximum dependency distance (max DD), as an indicator of source text difficulty to disfluencies in interpreting as an indicator of cognitive load. The authors compare sight interpreting of long max DD above the assumed working memory storage capacity of 7 chunks to sentences with short dependency distances and hypothesise that long max DD will cause more disfluencies in the sight interpreting process. As disfluencies, they include filled and silent pauses, repetitions, and repairs. Although their study is an experimental setup, dependency distances have also been used in a corpus-based approach to show implications for cognitive demands in different interpreting modes (Liang et al., 2017). Therefore, this approach is also listed here amongst other corpus-based studies. Jiang and Jiang (2020) show that sentence length, and long max DD, in particular, significantly influence the occurrence of disfluencies in the interpreting product with more disfluencies occurring with higher syntactic complexity. Their study further reveals that the type of syntactic relation for long distances plays a role

with subject-predicative, subject-verb, or nonfinite-verb as adverbial with long dependency distance resulting in increased cognitive load. Other types of relations with long dependency distances are not linked to an increase in occurrence of disfluencies. Contrastive differences between source and target languages and syntactic reshuffling may also play a role. The approach of using dependency distance measures to reflect cognitive ease is discussed in more detail in Section 2.3.2.

In summary, we can see that corpus-based approaches to investigate the interpreting process give indications as to what may cause peaks in cognitive load. These are particularly lexically dense input or complex syntax of the source (Plevoets and Defrancq, 2016, 2020), high source speech delivery rate (Plevoets and Defrancq, 2016), and low frequency words (Chmiel et al., 2023). Other factors such as high formulaicity (Plevoets and Defrancq, 2020) of the source (Chmiel et al., 2023) can reduce interpreters' load. These factors related to the source speech cannot be influenced by interpreters. They are, however, able to adapt their target speech production in ways that are beneficial for their own processing capacities, i.e. by producing formulaic sequences in the target text (Plevoets and Defrancq, 2018a, 2020) or using high-frequent cognates (Chmiel et al., 2023) as translation choice, leading to fewer disfluencies in the interpreting product. Further effects of the cognitively challenging production conditions in simultaneous interpreting are investigated in this thesis via lexical simplification operationalised through surprisal measures.

As the studies mentioned above do not find evidence of imported load or spillover effects (Plevoets and Defrancq, 2020; Chmiel et al., 2023; Shao and Chai, 2020), the sentence is a suitable unit of investigation for effects of cognitive processes in interpreting. This is particularly relevant for the approach described in Section 2.3.2, using Dependency Locality Theory as link to effects of cognitive processes, used in this thesis to investigate syntactic simplification.

2.3 Alternative approaches to approximate cognitive processing in interpreting product data

While the previous sections focussed on cognitive processes in interpreting and ways of capturing cognitive load via process and product data applied in simultaneous interpreting research, the following section gives an overview of methods of approximating cognitive load in interpreting used in this thesis, based on Information Theory and Dependency Locality Theory.

2.3.1 Information theory and interpreting research

Human language processing is assumed to be based on expectancy to a large extent. Above all, language users expect a message to be informative. For a conversation to be successful, speakers should follow the maxims on quality (being truthful), relation (being relevant), manner (avoiding ambiguity and obscurity, being brief and orderly) and quantity (not to convey more or less information than required) (Grice, 1975, p.47). The maxim on quantity or informativity is particularly relevant from an information-theoretic perspective.

Informativity of a message has been studied from various perspectives. Jaeger (2010) observes that language producers try to avoid peaks and troughs in information density and aim to distribute information evenly across the message (*uniform information density hypothesis*). Genzel and Charniak (2002) observe that recipients of a message expect a certain (high) level of informativity, and language producers will aim to be maximally informative without exceeding limits on entropy rates. Information rates that are too high or too low may even cause processing difficulties (Engelhardt et al., 2011; Kravtchenko and Demberg, 2015).

Expectancy in human language processing also refers to incremental processing. In online processing, such as spoken language comprehension, expectancy refers to the recipient of a message making predictions about what comes next (Crocker et al., 2016). Furthermore, in speech comprehension and production, we are primed by the general frequencies of linguistic units and rely on predictability in the linguistic and situational context for efficient comprehension and production (Hale, 2001; Levy and Jaeger, 2007; Levy, 2008). For language production, shorter and linguistically dense variants have been observed to be preferred in more predictive contexts, while longer and less dense linguistic variants tend to be used when the context is less predictive (Teich et al., 2020). This has been observed at various linguistic levels, for example, vowel length (Schulz et al., 2016), syntax (Reich, 2017; Delogu et al., 2017) or discourse relations (Rutherford et al., 2017). In some communicative events, redundancy within a message may be rational (Tourtouri et al., 2021).

Information theory based on Shannon (1948) models the communication process as an information channel with limited channel capacity with three basic components: *channel encoder, channel*, and *channel decoder*. "The channel encoder is the source of the message. It sends the code signal through the channel, the communicative medium. The channel may be noisy (adding distortion) or clean (without distortion). Finally, the channel decoder attempts to decipher the original message sent by the channel encoder." (Collins, 2014, p.652). Applied to the interpreting process, the use of the information channel is a two-stage process. In the first subprocess, the interpreter can be seen as the decoder of the original message sent by the original channel encoder in the source language. Then, they act as the channel encoder in a secondary communication channel, encoding information in the target language that was sent in the first communication channel, as rationally and truthfully to the source

message as possible. "Noise" can therefore affect the communication process twice. In the first subprocess, noise can be seen as anything that prevents the interpreter from fully decoding the source message, including actual acoustic noise, but also informativity above channel capacity, i.e. a source speech that is too informationally dense or covering information that the interpreter is not familiar with. If the source input is too fast and too dense (too much informativity at a rate that is too high), the interpreter will possibly not be able to decode the source message given the time restrictions during the interpreting process. In the second subprocess, noise can occur in the form of anything that affects the translation process as such. If lexical equivalents of source text items are not retrievable for the interpreter in the target language at the point in time required, *omission, generalisation, approximation*, or *summarising* can be seen as noise in the second subprocess, while the interpreter is still aiming to be as truthful as possible to the communicative goal. *Comprehension enhancing interpreting strategies* can be seen as aiming to prevent noise in the first subprocess, whereas *target text production interpreting strategies* relate to the second subprocess (Kalina (1998) cf. Section 2.2.2).

In the first subprocess, the interpreter has to rely on the encoder of the original message to encode the message rationally. In the second subprocess, the interpreter as encoder of the message in the target language is also excepted to encode the message in a rational way. Rational in this context can have two perspectives. As speakers aim to convey information to meet the audience's expectations, interpreters may design their output in a way that eases processing for their audience. However, in the context of the challenging production conditions in simultaneous interpreting, being rational in encoding a message and not exceeding channel capacity may mean that available cognitive resources have to be managed optimally. Optimal encoding in simultaneous interpreting may be different from optimal encoding in non mediated communication. Instead of designing the output to meet the audience's expectations, interpreters can be expected to aim at easing production. Interpreters may be assumed to produce expected output in order to ease target speech production and stay within the limited channel capacity, i.e. available cognitive resources.

Information theory (Shannon, 1948) provides a framework that captures language comprehension and production processes via probabilistic modelling of human language (Jaeger, 2010). Such language models can be used as the basis for product-oriented and process-oriented research in translation and interpreting studies, as well as to explain translational choices and their linguistic traces known as translationese (cf. Section 2.1) or interpretese (cf. Section 2.4). Information-theoretic measures that are applied to corpus data to reflect cognitive processes during language comprehension and production are *surprisal* – the predictability of a word in context – and *entropy* – the uncertainty of an outcome for a certain event.

The notion of *entropy* is related to various approaches in translation studies. *Relative entropy* or *Kullback-Leibler divergence (KLD)* is used to show distinctive characteristics of a language variety (original text or mediated language) when compared

to another language variety (Karakanta et al., 2021). *Translation entropy* (Moirón and Tiedemann, 2006) can be directly related to cognitive effort in language comprehension. It is the likelihood of selecting one translation solution from a search space of possible translations (Teich et al., 2020). It can be used to quantify translation difficulty (Martínez and Teich, 2017; Carl et al., 2016). For default translations, i.e. source items that are always translated with the same target item such as *European Union - Europäische Union*, translation entropy is very low. If no default translation is available and the translations space for possible translations solution is large, translation entropy and thus the cognitive effort required to select this translation solution is high (Wei, 2022). The implementation of the translation entropy approach generally requires parallel corpus data (source and target texts) with multiple translations for each source text or source text item.

Schaeffer et al. (2016) and Carl and Schaeffer (2017) demonstrate that translation entropy correlates with word fixation duration in reading times. They observe a short first fixation on a word, indicating easy retrieval, for low translation entropy items, while higher processing effort or difficulty with longer fixation is linked to higher translation entropy (Schaeffer et al., 2016; Carl and Schaeffer, 2017). With a focus on discourse relations, Pollkläsener et al. (2024) investigate connectives and their translation equivalents. Using translation entropy the authors show that cognitive load in the translation process of connectives experienced by translators and interpreters differs as the same connectives have more translation equivalents and higher translation entropy in translations than interpreting. They link their result to simplification and explicitation effects. Yung et al. (2023) measure translation entropy of the corresponding connectives in parallel texts using the sense distribution for each connective. Using this method, they observe explicitation in translations by use of more specific connectives.

This thesis uses two information-theoretic measures. *Relative entropy* or *Kullback-Leibler divergence (KLD)* is used to detect distinctive properties of interpreted language, and *surprisal* aims to investigate simplification in interpreter output. Therefore, related work for these two methods is given in the next sections.

Relative entropy (Kullback-Leibler Divergence, KLD)

Relative entropy can be applied to show how one language variety differs from another language variety by measuring the number of additional bits that are needed to model the language variety when a non-optimal model is used. KLD is an asymmetric variant of relative entropy, giving different results for language variety A compared to variety B than for language variety B compared to variety A. Fankhauser et al. (2014) apply this approach to the Brown corpus, showing distinctive differences in English text registers visualised via word clouds. The results given in the approach are based on the most distinctive words (obtained via relative entropy) combined with results for the most frequent words. The same approach is applied by Degaetano-Ortlieb and

Teich (2018) to characterise the course of diachronic change and the features that contribute to this change in late modern English scientific writing or by (Karakanta et al., 2021) to detect specific properties of translated language.

In this context, further related measures should be mentioned. The symmetric variant of relative entropy, Jansen-Shannon Divergence, can also be used to detect linguistic features that distinguish a specific language variety from another. Klingenstein et al. (2014) use this symmetric type of relative entropy to identify speaking styles in criminal trials. Furthermore, *relative entropy* has also been shown to be useful in improving efficiency in statistical machine translation. Kim et al. (2017) apply KLD to improve the efficiency in pruning, that is, to reduce the number of parameters or computational resources that are used in a language model. KLD and variants absolute KLD as well as symmetrised KLD have been shown to be useful in detecting redundancy through mutual information for zero and negative divergence results (Kim et al., 2017).

Compared to more traditional, frequency-based methods in corpus linguistics, the KLD approach has the advantage of being data-driven and no prior selection of features is required. It also includes built-in significance testing, which provides a more objective procedure and makes interpretation of results easier. Only significant features are inherently considered. In this thesis, KLD is used to detect characteristic properties of interpreted language (*interpretese*) when compared to spoken originals of the same language. It is applied as described in Karakanta et al. (2021), who base their approach on Fankhauser et al.'s (2014) implementation via word cloud visualisations. Section 4.1 explains the method of relative entropy applied in this thesis to interpreting data and original spoken production.

Surprisal

Surprisal, also known as *Shannon information, information density* or *self-information*, reflects the intuition that unpredictable items would carry more information than predictable ones (Collins, 2014, p.652). It is derived as the negative logarithm of probability of a given unit (e.g. word) in its context (e.g. the preceding n words) (Shannon, 1948). In the linguistic context, words with low surprisal are highly expected in the given context, whereas high surprisal words are low in probability to occur in their context and therefore convey more information[1].

Surprisal has been shown to be a good quantification of cognitive load in human language processing as it correlates with several behavioural and neurophysiological indices such as reading times, pupil dilation, response latencies, or event-related potentials (ERPs) (Crocker et al., 2016; Hale, 2001; Levy, 2008; Levy and Gibson, 2013). In theoretical terms, surprisal is seen as incremental sentence comprehension

[1]More information on the calculation of surprisal for the data used in this thesis is given in the corpus section (Section 3.1.1) and the methods section of the related study on lexical simplification (Section 5.2.1).

as "step-by-step disconfirmation of possible phrase-structural analyses for the sentence, which means that cognitive load can be interpreted as the combined difficulty of disconfirming the disconfirmable structures at a particular point of the sentence (i.e. at a given word)" (Wei, 2022, p.78). Highly unexpected input (high surprisal) may cause a shift in the allocation of cognitive resources during comprehension to find possible alternatives due to uncertainty (Levy and Gibson, 2013).

Probability estimations to investigate human language processing can be obtained using different approaches. The approach taken in this thesis is to calculate the probability estimations retrieved from n-gram models calculated from the interpreting corpus data investigated in this thesis (cf. Section 5.2.1). Alternatively, experimentally orientated studies in psycholinguistics would commonly use cloze probabilities to estimate surprisal, while more recently large language models (LLMs) have been shown to be suitable for retrieving probability estimations (Michaelov et al., 2023; Oh and Schuler, 2023).

In translation and interpreting studies, *surprisal* is used to detect specific characteristics of translated or interpreted texts, as well as for research on the translation process. Tapping into one aspect of the translations process, Martínez and Teich (2017) use surprisal to assess acceptability of a learner's translation solution by assuming surprisal values in learner translation that are closer to surprisal in professional translations are linked to higher acceptability of the learner's translation. Translating specific terminology requires routine translation that generally results in very expected translation solutions. The authors show that for specific terms where professionals use the expected routine translations with very low surprisal, translation students use greater variation of translation solutions and a lower degree of routinisation. This indicates greater confidence in decision-making and a higher degree of automation within professional translators (Martínez and Teich, 2017). A high degree of automation is also believed to be essential during simultaneous interpreting (cf. Section 2.2.2 on interpreting strategies), and we therefore assume routinised and low surprisal translation solutions in interpreters' target production for expected terms in the source speech.

More recently, information-theoretic measures have been combined with translation process data. Several studies use the CRITT TPR-DB database (Carl et al., 2016) to investigate the link of surprisal and translation difficulty via process data such as reading times of sources or target text items, or translation duration, obtained via eye-tracking and keystroke logging during the translation process. Schaeffer et al. (2019a) even state a "predictive turn" in translation process research involving computational modelling of the human translation process combined with behavioural measures of the translation process.

Carl (2021) use surprisal retrieved from their CRITT TPR-DB database (Carl et al., 2016) as a predictor of translation difficulty. A similar approach is taken by Wei (2022) who use surprisal and entropy to predict the target text production time

as well as the source and target text reading times. Their CRITT TPR-DB dataset (Carl et al., 2016) includes six English source texts that are translated into various languages. For target text production time, surprisal was found to be the strongest predictor in their model, which also includes entropy as the main predictor variable. For the source text reading time, entropy was found to be more reliable than surprisal, and for the target text reading time, no large differences in the strength of prediction was found for entropy vs. surprisal.

Lim et al. (2023) aim to confirm that current NLP models such as LMs and NMT align with human language usage to some extent and can be used to predict language processing complexity. The authors analyse data based on 13 language pairs and hundreds of human translators from the CRITT TPR-DB database (Carl et al., 2016). They use monolingual surprisal derived from a monolingual language model (LM), as well as translation surprisal obtained from the distribution of a neural machine translation (NMT) model to predict reading times (of source and target text), as well as target text production duration. The NMT model is conditioned on the token sequence of the source text and previously translated tokens in the target text. They find monolingual surprisal to be the best predictor of source text reading time and target production duration, but not of target text reading time. For the target side, they find that surprisal derived from the neural machine translation model is a more consistent predictor of translation difficulty (target reading time and production duration). The authors conclude that state-of-the-art models are suitable to capture important aspects of translation difficulty (Lim et al., 2023).

In addition to its usability in translation process research, surprisal is also applied for translationese detection. Rubino et al. (2016) use surprisal features in a translationese classification task. They show that surprisal features on word, part-of-speech and syntax level are suitable to distinguish translations from non-translations. Furthermore, the notion of surprisal is also used in translation and interpreting studies to link phenomena that occur in translated or interpreted language to translationese / interpretese trends. Lapshinova-Koltunski et al. (2022) use surprisal of discourse connectives and link it to well-known translationese phenomena (cf. Section 2.1.2) of explicitation and implicitation. Using the corpus data and modelling approach used in this thesis (EPIC-UdS, Przybyl et al. (2022b)) combined with a comparable written dataset (Europarl-UdS, Karakanta et al. (2018)), Lapshinova-Koltunski et al. (2022) investigate discourse connectives and their translations from English into German and German into English in written and spoken translation. The authors assume that connectives with low surprisal in the source texts would require low cognitive effort for a translator or an interpreter as a recipient. At the same time, target side surprisal indicates cognitive effort of a recipient of the translated or interpreted texts (reader or listener). Comparing distribution of equivalence, implicitation and explicitation strategies, as well as surprisal of the same connectives in comparable source and target contexts, the authors show that equivalence, i.e. that a connective in the source is also transferred to the target, and implicitation, when a connective

is omitted or replaced with a more general connective, is used more frequently in interpreting and may facilitate cognitive processing in high-time-pressure situations. In the written translation mode, the authors show that explicitation may provide an advantage in cognitive processing effort for the recipient, although adding a connective that is not present in the source or using a more specific connective requires more effort for the translator / interpreter. Explicitation in such cases is seen as audience design, occurring in translations rather than in interpreting (Lapshinova-Koltunski et al., 2022).

Touching on the translation process and translationese / interpretese phenomena, Kunilovskaya et al. (2023b) investigate how information density of a source text affects informativity of the translated or interpreted target text. Using the same dataset as Lapshinova-Koltunski et al. (2022) they use a slightly different modelling approach to retrieve surprisal values: Kunilovskaya et al. (2023b) use a strictly balanced dataset of original and mediated (translated or interpreted) data and retrieve lemma-based surprisal values. Average segment surprisal is used to correlate informativity of source and target. They find that generally, information density of the target text is conditioned by informativity of the source text: low information input leads to low information output and informationally dense input leads to informationally dense output. For the same level of surprisal in the source, interpreters produce lower surprisal output than translators (Figure 2.4). This can be interpreted as interpreters using their available cognitive resources efficiently to cope with the challenging production conditions in simultaneous interpreting.

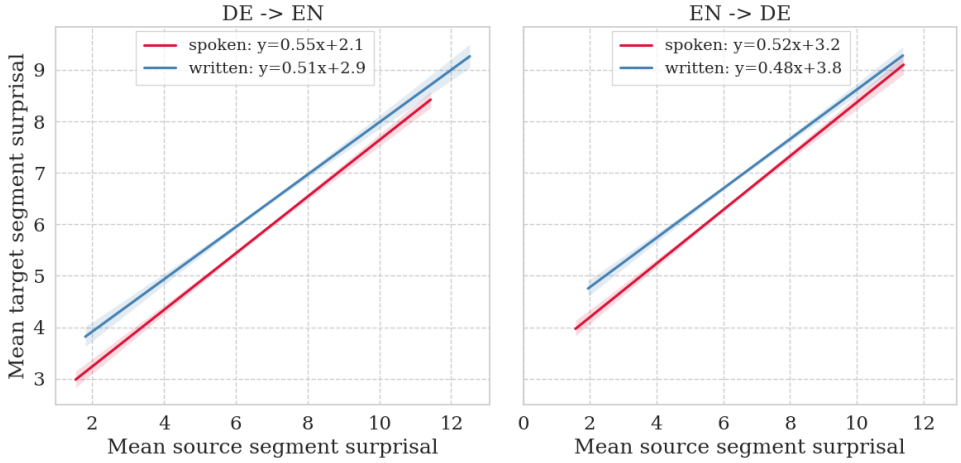


Figure 2.4: Linear regression based on AvS of aligned source and target segments by translation direction (DE-EN, EN-DE) and mediation type (spoken/interpreting, written/translation) (Kunilovskaya et al., 2023b, p.612).

Kunilovskaya et al. (2023b) also use surprisal to investigate the impact of challenging production conditions that may influence interpreters' ability to fully transfer the source text informativity to the target text. Prepared and read-out source speeches

are assumed to be more informationally dense and more difficult to interpret. However, the authors find no evidence of these assumed more challenging conditions, leading to lower surprisal in the target. On the contrary, interpreters seem to be able to encode the same level of information regardless of the type of source delivery (Kunilovskaya et al., 2023b, p.613). Delivery rate, on the other hand, as an explanatory variable for the target speech information density, gives the result of higher source speed rates (and therefore less time for interpreters to process the input and retrieve and produce the target output), leading to lower informativity in the target (Kunilovskaya et al., 2023b, p.614).

Taking a monolingual, comparative perspective, Kunilovskaya et al. (2023b) also investigate the simplification hypothesis as specific translationese / interpretese phenomenon. They compare average segment surprisal of original texts in one language (either German or English) to interpreted texts into the same language. Surprisal approximates simplification by assuming that lower surprisal and therefore more expected words in context are easier to process for the recipient. Kunilovskaya et al. (2023b) show simplification indicated by lower information density in interpreting compared to original spoken texts in the same language (cf. Figure 2.5). However, for the written mode, Kunilovskaya et al. (2023b) show that this translationese trend generally depends on the languages studied (cf. Mauranen, 2007), with simplification only observed for translations into English in their study, and a reversed trend for translations into German (Kunilovskaya et al., 2023b).

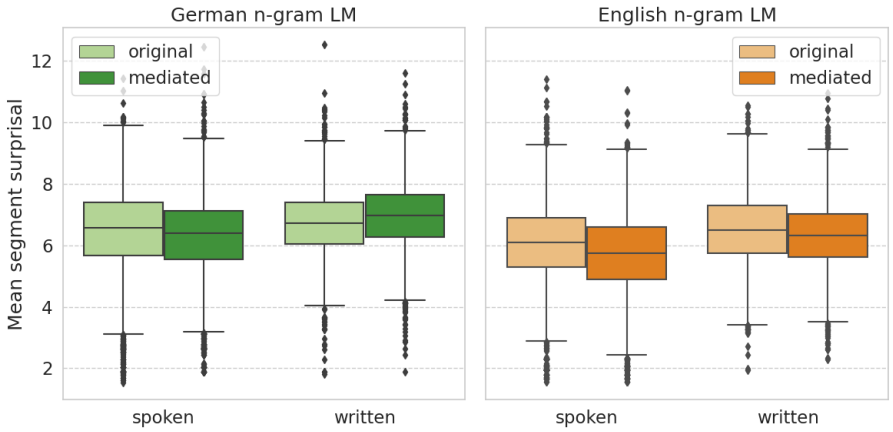


Figure 2.5: Average surprisal of segments across the subcorpora (Kunilovskaya et al., 2023b, p.613).

In summary, information-theoretic measures and surprisal in particular have been found to reflect the translation process via routine translation solutions (Martínez and Teich, 2017) or to predict translation production times (Wei, 2022; Lim et al., 2023). Furthermore, surprisal can be used for automatic translationese detection (Rubino et al., 2016) or be linked to individual translationese tendencies such as simplification

and explicitation (Kunilovskaya et al., 2023b; Lapshinova-Koltunski et al., 2022).

This thesis aims at giving a clearer description of the interpreting product and therefore uses relative entropy to detect the specific characteristic of interpretese. Particularly in the context of the processing constraints in simultaneous interpreting, surprisal will be used to investigate the simplification hypothesis, assuming that interpreters will produce more expected output than comparable source speakers. Contrastive studies for English and German, the language pair investigated in this thesis, show that information density is distributed differently in both languages (Doherty, 2006; Fabricius-Hansen, 1996). This may affect the distribution of information density in interpreted language. It is therefore important not only to take a comparable approach to detect simplification in the interpreted speech but also to look into potential triggers that may cause peaks and troughs in information density by investigating high and low surprisal items.

Using surprisal to investigate lexical processing is one way to approach cognitive processes. Several studies have combined this approach with another approach that considers syntactic integration and storage costs in speech processing using Dependency Locality Theory (DLT) (e.g. Collins, 2014; Dammalapati et al., 2021; Rajkumar et al., 2016; Balling and Kizach, 2017). Some authors find the locality effect to be a more robust predictor of sentence processing than lexical surprisal (Balling and Kizach, 2017). Others view Information Theory and Dependency Locality Theory as two complementary models for cognitive processing (Collins, 2014) and show that lexical surprisal and DLT integration and storage costs give good results in explaining spoken language production (Dammalapati et al., 2021). Futrell (2019) shows the direct link between information theory and dependency locality by proposing that information locality subsumes the principle of DLT. He proposes that efficient languages should minimise the linear distance between elements carrying high mutual information.

2.3.2 Dependency Locality Theory and interpreting research

As laid out in Section 2.2.1, memory plays an important role in the interpreting process. Dependency Locality Theory, (DLT) (Gibson, 1998, 2000) and particularly Dependency Length Minimisation (DLM) provide a direct link to optimising memory utilisation in language processing. This section explains the concept behind DLT as a further corpus-based approach to cognitive processes in language processing. It explains its relevance for language processing and production before introducing general principles on DLM and related studies, particularly in the multilingual context. The section ends with an overview of DLM applications in interpreting research.

Dependency Locality Theory

The second framework to account for cognitive processing in interpreting, and memory in particular, used in this thesis is Dependency Locality Theory, DLT (Gibson, 1998, 2000). It was originally proposed as a theory of resources limitations, which is an aspect that is crucial during the interpreting process. It gives a measure of the syntactic complexity of a sentence by assigning processing costs via storage and integration costs to all elements of a sentence. In this context, storage costs are linked to memory with shorter dependencies leading to optimised memory utilisation.

The concept behind this is as follows. Gibson (2000) assumes that for sentence processing by the human brain, words are processed one by one. For sentence processing, two components are believed to require cognitive resources: (a) Integration costs for new words to be integrated into an existing structure (such as heads and their dependents) and (b) memory costs for storing incomplete dependencies. Structural integration costs are related to locality (or shorter dependencies), with greater locality relating to lower integration costs, while greater distance is reflected in higher integration costs (Gibson, 2000, p.102).

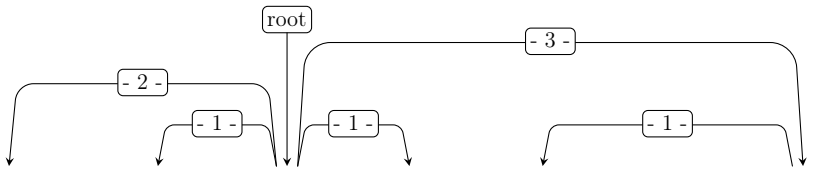
The basic principle of DLT is that each element in a sentence depends on another element in the sentence: head and dependent. Every dependent is linked to exactly one head. The only exception is the root of the sentences, which serves as the head of the entire sentence (Gildea and Temperley, 2010). Consider the following examples. Sentence A (example 7) is a corpus example of EPIC-UdS, produced by an interpreter. Sentence B (example 8) is a grammatically correct alternative. The two examples above differ only in the position of *sein*, with the interpreter producing the main verb as early as possible and, therefore, reducing the distance between head (in this case *Teil*) and dependent (*sein).*

(7) [A] es muss *Teil sein* der Unternehmensstrategie.
(EN: it must part be of the corporate strategy)

(8) [B] es muss *Teil* der Unternehmensstrategie *sein.*
(EN: it must part of the corporate strategy be)

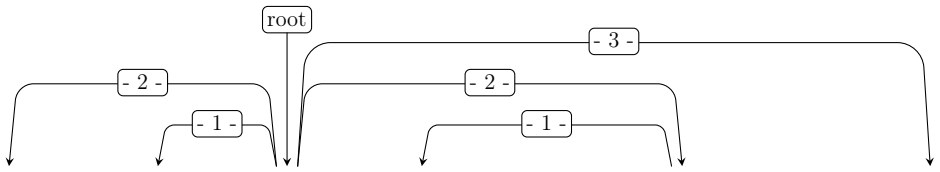
The dependencies within a sentence can be better visualised via arrows, e.g. in Figure 2.6. The head and dependents can be closer (greater locality, Figure 2.6) or more distant (greater length, Figure 2.7) as indicated by the number above the arrow. For structural integration, higher costs occur for dependencies with greater distance.

When perceiving a word w, maximal projections of syntactic predictions are created for all syntactic categories that can possibly follow w as the next word in a grammatical sentence. For *structural integration*, the syntactic category is matched with the predictions made for the syntactic structure already built. Other factors such as frequency of the relevant lexical entry or new elements to the discourse, requiring



es	*muss*	*Teil*	*sein*	*der*	*Unternehmensstrategie*
(EN: it	must	part	be	of the	corporate strategy)

Figure 2.6: A - Corpus example, dependency annotated.



es	*muss*	*Teil*	*der*	*Unternehmensstrategie*	*sein*
(EN: it	must	part	of the	corporate strategy	be)

Figure 2.7: B - Alternative to corpus example, dependency annotated.

more resources to process than already introduced elements, also play a role in discourse processing. The ease of integration may also depend on whether an element is focused or unfocused, new or previously introduced, etc. (Warren and Gibson, 2002).

Altogether, "the difficulty of the structural integration depends on the complexity of the structural integrations in the interim, as well as on the discourse integrations and the plausibility evaluations in the interim" (Gibson, 2000, p.103). Although structural difficulty and contextual plausibility also play a role for structural integration, they are ignored in the DLT approach.

Figure 2.6 (and sentence 9) and Figure 2.7 (and sentence 10) display the example sentences from above. They indicate the head and dependent relationship relevant for structural integration via arrows above the words, as well as the distances as numbers above the arrows or in brackets within the sentences. As the sentences only differ in the position of *sein*, structural integration costs can be compared directly via dependency distances. The largest distance is 3 words apart from the root in both sentences. However, when summing the distances for both sentences, sentence A (Figure 2.6) shows lower integration costs with summed distances of 8 vs. summed distances of 9 in the alternative example (sentence B, Figure 2.7).

(9) [A] es (2) muss (1) Teil (root) *sein* (1) der (1) Unternehmensstrategie (3)

(10) [B] es (2) muss (1) Teil (root) der (1) Unternehmensstrategie (2) *sein* (3)

DLT is a simplified account of sentence processing at various levels. Nevertheless, it accounts for more complex sentence structures, considering that structural integration is not a linear process. Hence, integrating nested structures or multiple integrations are seen to increase processing costs simply by taking into account that many long-distance structural integration steps have to take place at the same point when processing nested structures (Gibson, 2000). Gibson (2000) shows that DLT predictions for complex structures are in line with reading times measures in a self-paced reading experiment (Gibson and Ko (1998) cited in Gibson, 2000). Specifically, for relative clauses with subject and object extraction Gibson (2000) shows that subject-extracted relative clauses are easier to process, with shorter distances for subject and dependent relative pronoun (they are usually adjacent) than for object-extracted relative clauses with longer dependencies. Furthermore, Gibson (1998) also provides cross-linguistic support for the DLT, showing its application in different languages, such as English, Japanese, Dutch, German, Spanish, and Finnish.

In addition to integration costs, the second cost-effective process in sentence processing that requires cognitive resources is storing incomplete syntactic structures or dependencies. Gibson (2000) counts storage cost in memory unit (MU). A MU relates to the minimal number of syntactic head categories that a sentence requires to be grammatically complete. In other words, the minimal number of words the sentences requires to complete all open dependencies.

For the two example sentences above, this can be stated as follows (MUs in brackets):

(11) [A] es (1) muss (1) Teil (1) *sein* (0) der (1) Unternehmensstrategie (0)

(12) [B] es (1) muss (1) Teil (1) der (2) Unternehmensstrategie (1) *sein* (0)

When processing the sentence initial word "es", at least a predicate must follow. Therefore, in both examples (11 and 12) the amount of required MUs at this point of sentence processing is 1. The auxiliary verb "muss" that follows requires at least a main verb (at least one more head to follow, MU = 1). "Teil" follows, therefore, a main verb is still missing (MU = 1). In example 11 [A], the dependency is closed by the main verb (MU = 0). The following "der" requires one more head (MU = 1). With "Unternehmenstrategie", all dependencies are closed again (MU = 0). Example 12 [B] is more complex from a storage point of view as instead of closing a dependency, "der" opens a new open relation (MU = 2). This is closed with the head "Unternehmenstrategie" – there is only one open relation at this point of sentence processing. At the end of this sentence, the main verb "sein" closes all open relations, bringing the MU to 0. Closing the dependency earlier in example 11 [A] is beneficial from a memory perspective.

Short-term memory and memory capacity limitations are crucial in simultaneous interpreting, as interpreters need to store incoming speech signals before and while producing the corresponding target segment. Memory resources need to be efficiently managed. This is observed by Collard et al. (2019), who find a tendency of simultaneous interpreters to close dependency relations earlier than original speakers in the same language. The authors investigate subordinate clauses in Dutch and German which generally follow a Subject-Object-Verb structure. The verb is generally produced last in a Dutch or German subordinate clause. It is possible, however, to produce the verb earlier and extrapose items that otherwise would be produced before the verb, as in example 13. With the production of the main verb "haben", MU in Gibson's (2000) terms is reduced before increasing again with the extraposed information that follows. Such an early production of the head-dependent relation is an exception in both languages, but happens statistically more frequently in interpretations into German compared to original spoken production in German (Collard et al., 2019).

(13) ...die vor allem zu tun *haben* mit der Einrichtung einer europäischen Aufsichtsbehörde oder auch mit der europäischen Registrierung und Zulassung euh euh von euh Wertpapierverwaltungsgesellschaften
(EN: ...which mainly have to do with establishing a European supervisory body or also with European registration and certification of portfolio management companies)...

An early production of a closing dependency relation might be beneficial in terms of interpreter's working memory management. However, particularly in a multilingual setting, other factors can also influence dependency distances. Regarding comprehension difficulty at a word in a sentence, Gibson claims that this depends on the combination of the following factors (Gibson, 2000, p.105):

1 the frequency of the lexical item being integrated

2 the structural integration cost at that word

3 the storage cost at that word

4 the contextual plausibility of the resulting structure

5 the discourse complexity of the resulting structure

DLT accounts for increasing structural integration costs with distance of attachment (2) and increasing memory costs with predicting syntactic categories (3). Even though (1), (4) and (5) are neglected in DLT, its validity has been shown cross-lingually for complex structures, such as multiple embedded structures, nested clauses, or ambiguous structures (Gibson, 2000) via reading time experiments. Furthermore, it is a good operationalisation for measuring syntactic complexity, regardless of plausibility and frequency effect (Gibson, 1998, 2000).

Dependency Length Minimisation

Gibson (1998, 2000) lays the theoretical foundation that structures with longer dependencies are supposed to be more difficult to process. A large and diverse body of research supports this theory by showing that natural languages tend to have a preference for shorter dependencies and propose a preference for Dependency Length Minimisation (DLM) (Gildea and Temperley, 2010; Temperley, 2007). This applies to language comprehension (Gibson, 1998, 2000; Diessel, 2005; Liu, 2008) as well as to language production and is especially related to syntactic choice, where the language producer has different options to communicate the same message (Hawkins, 1995, 2004; Temperley, 2007, 2008; Liu et al., 2017; Ferreira and Dell, 2000; Arnold et al., 2000, amongst others). Long dependency distances can cause comprehension difficulties, as measured in longer reading times (Grodner and Gibson, 2005; Bartek et al., 2011). Although DLM relates to language comprehension and production, greater locality seems to be more beneficial for production than for comprehension (Konieczny, 2000). Tendencies to minimise dependency distances can be observed across many different natural and artificial random languages (Liu, 2008; Futrell et al., 2015; Gibson et al., 2019). Futrell et al. (2015) even suggest DLM to be "a universal quantitative property of human languages" (p.1). Dependency distance minimisation tendencies are also assumed to have influenced the evolution of natural language grammars (Gildea and Temperley, 2010). Furthermore, DLM can also be observed across difference genres (Wang and Liu, 2017).

Grammars generally allow for a certain amount of choice to structure syntactic components. Some theoretical principles are laid out by Temperley (2007) that, if acceptable in a language, would allow for minimisation of dependency length. Theoretically, dependency length can be minimised if dependents are always either on the left or on the right of their head (always left-branching or always right-branching) (Temperley, 2007, p.304). General tendencies of languages being primarily left-branching or right-branching can be observed, with English falling into the category of mainly right-branching (Temperley, 2007). "In a primarily right-branching language, the left-branching constituents should be short" (Temperley, 2007, p.306). If a head has several dependents (i.e. the root of a sentence), dependency length can be minimised when the shorter of two dependent phrases is placed closer to its head (Hawkins, 1995, 2004; Temperley, 2007; Gibson, 1998). Furthermore, placing dependent phrases of the same length on each side of the head (referred to as "opposite-branching") is also beneficial to reduce dependency length (Temperley, 2008). A tendency of long before short dependency seems to be preferred in SOV languages whereas short before long preferences exists for SVO languages (Liu et al., 2017).

An approach to investigate DLM in natural languages is to compare original corpus data with actual dependency length to an optimised ordering, using actual grammar rules for that language, but ordering dependents in a way that minimises dependency length. The results of these optimised orderings are also compared to completely ran-

domly ordered dependents and their overall distances (Gildea and Temperley, 2010). In a study with 37 languages, Futrell et al. (2015) find shorter dependency length in all natural languages compared to random ordering. This effect is significant for sentences longer than 12 words and is more pronounced for long sentences. Observed dependency length is still longer in their corpus data across all languages and language families compared to grammatically acceptable optimised dependency distance, which points to other factors besides dependency minimisation also influencing language use. An example is a linearisation effect where utterances can be understood incrementally with the comprehender being able to quickly identify the syntactic and semantic properties of each word (Hawkins, 2014). Futrell et al. (2015) generally find more minimisation for head-initial languages than for head-final languages and link this observation to richer morphology such as case marking on dependents in head-final languages such as German. Furthermore, non-canonical word order choices can contribute to dependency minimisation trends (Yu et al., 2019). Since the Universal Dependencies framework and available treebanks have been published (Nivre et al., 2016), more multilingual studies of DLM have been possible with consistently annotated dependency structures (Dyer, 2023). The Universal Dependency framework is also used for the annotation of dependency length in this thesis.

DLM in a contrastive perspective (English / German)

As this thesis investigates English and German data, specific dependency structure observations for these two languages are considered in more detail. As for English, syntactic ordering is mostly governed by rules. Nevertheless, some flexibility in ordering dependents of a verb remains. For example, a prepositional phrase usually follows a direct object noun phrase. However, if the noun phase is long, the order tends to be reversed (referred to as "heavy-NP shift") (Hawkins, 1995). Such an ordering keeps the shorter dependent phrase closer to the verb and, therefore, reduces dependency length. Furthermore, English tends to be primarily right-branching, but in certain cases, grammatical rules allow for left-branching structures such as subject NPs (left-branching) that are generally shorter than right-branching object NPs (Temperley, 2007). With a general SVO sentence structure and the head in an NP generally being close to the left of the NP (with the verb as root therefore head for both subject and object phrases), the SVO ordering in this case favours minimisation of dependency distances (Temperley, 2007). In certain cases, other orderings that allow for shorter dependency length are possible, such as subject-verb inversion (e.g. in quotation constructions), and for such verb-subject constructions, the subject NPs (in this case right-branching) are found to be much longer than in subject-verb constructions (left-branching), thus confirming a DLM trend in English (Temperley, 2007). A further example for dependency length minimisation orderings are right-branching, postmodifying adverbial clauses in English that tend to be significantly longer than left-branching, premodifying ones. These observations confirm the general minimisa-

tion trend of English as primarily right-branching language showing a preference of left-branching to be short (Temperley, 2007). Giving a last example, the tendency found in English to use opposite-branching for one-word phrases is also favourable for minimising dependency length (Temperley, 2008).

Corpus studies find a strong tendency of dependency length minimisation in written English (Gildea and Temperley, 2010; Futrell et al., 2015; Dyer, 2023, amongst others). Original English corpus data are much closer to an ordering of constituents that would be optimal for DLM than random ordering and this can be attributed at least partly to syntactic choices that reduce dependency distances (Gildea and Temperley, 2010). If there are syntactic alternative ordering choices, dependency length minimisation tendencies can be particularly observed for long and moderate length dependencies (Rajkumar et al., 2016). For short dependencies, a facilitating effect is also observed for processing short subject-verb dependencies in English main clauses (Bartek et al., 2011). Due to the ordering choices observed in corpus data, some scholars refer to English as using a head-medial, balanced distribution, where the head of a phrase has an almost equal number of dependents on each side (Xu and Liu, 2023).

Compared to English, German is often described as a language with a freer word order (Gildea and Temperley, 2010; Kempen and Harbusch, 2008), which generally goes hand in hand with rich case marking / morphology (Gibson et al., 2019). An example for free word order are arguments of the verb (subject, direct, and/or indirect object), which are highly variable in their position in the German sentence (Kempen and Harbusch, 2008). Theoretically, the dependency length can be minimised in German by using the following ordering choices: In the case of relative clauses, dependency length tends to be shorter if a long relative clause is extraposed of a short main clause compared to the non-extraposed variant of the same sentence (Konieczny, 2000). Long relative clauses thus favour extraposition (Gildea and Temperley, 2010). Similarly, some long and complex verb complements can be extraposed in German, which minimised dependency length compared to the non-extraposed variant (Hawkins, 1995). In most languages, dependencies do not cross each other nor cross the root of a sentence (Ferrer-i-Cancho, 2006). However, in German, crossing dependencies are possible and also observed in corpus data, such as in the context of extraposition of elements of a subordinate clause (Gildea and Temperley, 2010; Brants et al., 2004). Such extraposition may reduce dependency distances and may be beneficial at least concerning the memory costs (i.e. MUs in Gibson's (2000) terms). For syntactic orderings with the verb in final position as may occur in German subordinate clauses, it is generally observed that long constituents are more likely to precede short ones in the preverbal position (Hawkins, 1995). Furthermore, ease of processing shorter dependencies can also be found in subject-extracted relative clauses (Levy and Keller, 2013).

In contrast to the observed DLM trend in English, German original corpus data do

not show such clear signs of reduced dependency length compared to an optimal or random ordering of sentence constituents (Gildea and Temperley, 2010) and overall weak evidence for DLM in German (Park and Levy, 2009). German actual dependency length is only slightly lower than the random ordering (Gildea and Temperley, 2010). Gildea and Temperley (2010) propose this to be due to phenomena that are not captured by grammatical rules but rather by syntactic choice. Furthermore, they propose the idea that languages with greater use of syntactic inflections such as German with case and gender markings may also allow longer dependencies (Gildea and Temperley, 2010, p.305). Generally, head-final languages show less dependency length minimisation trends than head-initial languages (Futrell et al., 2015; Jing et al., 2022; Dyer, 2023). At the same time, long dependencies in German with its case markings and rich morphology have been found to be easier to process than long dependencies in English (Konieczny, 2000).

Longer dependencies in German may particularly be linked to the final position in clauses such as participles and finite verbs in dependent clauses (Gildea and Temperley, 2010). The distance between final participles and preceding auxiliary verbs can be large with many other constituents between head and dependent. The verb final position in the clause leads to crowding of constituents on the left side of the (head) verb. Finite verbs generally have long dependencies to both their heads and dependents as opposite branching of dependents is not possible. At the same time, finite verbs have the maximum distance to their head (Gildea and Temperley, 2010). Although there is a general possibility to extrapose elements of the dependent clause, this syntactic choice may only be preferred in certain constrained production settings (Collard et al., 2019). Furthermore, auxiliaries and participles in English are mostly adjacent (with a distance distance of only 1) (Gildea and Temperley, 2010). In German, fewer adjacent dependencies are observed than in English (Liu, 2008).

DLM is found in English to a much higher degree than in German (Gildea and Temperley, 2010; Futrell et al., 2015; Dyer, 2023). The actual dependency length in English is much closer to an optimal ordering than a random arrangement, whereas actual dependency length in German is closer to a random than an optimal ordering (Gildea and Temperley, 2010). In addition to the processing advantages linked to shorter dependencies, other factors may also influence the ordering of syntactic constituents, such as animacy, with animate nouns tending to precede inanimate nouns as animate objects are supposed to be more accessible (Kathryn Bock and Warren, 1985), information status, with given elements tending to precede new elements (Arnold et al., 2000), or semantic connectedness (Hawkins, 2004). In general, dependency minimisation in natural languages may be seen as a general trend rather than an absolute rule (Liu et al., 2017). As human language production is subject to multiple constraints in some cases dependency distance minimisation may have to be sacrificed for the benefit of reliable and effective communication "it may exploit other strategies to counter-balance the memory burden imposed by long dependency

distance and thus give rise to some unique linguistic patterns with long dependencies" (Liu et al., 2017, p.1186). Such strategies used to counter-balance the memory burden may be the use of high-frequency functional elements (Liu et al., 2017), e.g. the use of a complementizer *that* is more likely in long dependencies than short dependencies (Jaeger, 2010). Optional function words or other dependents may be omitted to shorten dependencies (Jaeger, 2010; Liu et al., 2017). Furthermore, frequency and conditional probability may be linked to easing the processing difficulty of long dependencies (Liu et al., 2017). These factors generally refer to monolingual language processing. A multilingual setting such as translation and interpreting in particular may also be influenced by properties of the source language in general, as well as structural properties of the source text.

Most of the studies mentioned above refer to written production. However, there may be differences between spoken and written production as to whether syntactic choices are made due to production or comprehension pressures (Temperley, 2007), or rather both (Rajkumar et al., 2016). Generally, Gildea and Temperley (2007) observe similar dependency minimisation patterns for written and spoken language, but there may be differences as to whether these patterns can be linked to production or comprehension facilitation. Ferreira and Dell (2000) and Arnold et al. (2000) both focus on the spoken mode and argue that certain syntactic choices indeed facilitate rapid and fluent speech production. With more time, the possibility of rewriting and optimising the production, syntactic choice may be linked to facilitate comprehension in the written mode (Temperley, 2007). For spoken language, it seems unlikely that speakers actively seek to facilitate comprehension and adjust their speech to the needs of the listeners but rather pursue "their own communication goals or according to availability-based production accounts, they might be realising the most readily available constituents" (Rajkumar et al., 2016, p.225). In fact, Dammalapati et al. (2021) show this production focus in spoken language use by linking DLT-based storage and integration costs to increased cognitive load in speech production indicated by the occurrence of disfluencies. This particularly shows the production focus in spoken language. Therefore, speakers may realise the most accessible elements and memory-based effects such as dependency length minimisation rather seem to be the result of the production process and not primarily a means of audience design (Rajkumar et al., 2016).

DLM in interpreting research

As elaborated above, locality phenomena can be linked to working memory limitations during language processing. For spoken language, this is particularly related to language production (Konieczny, 2000; Wasow, 1997; Rajkumar et al., 2016), but is also relevant for language comprehension Gibson (1998, 2000). The interpreting process covers both aspects with dependency measures in the source text relating

to interpreters comprehension difficulty, and dependency locality in the target text relating to target text production. With interpreting as a particularly cognitively challenging task, it can be assumed that interpreters may try to minimise dependency distances to facilitate target text production. Interpreting corpus data such as EPIC-UdS (Przybyl et al., 2022b) reflect the language production process for source speakers and interpreters because no corrections are made to the speaker's nor interpreter's output.

The section on corpus-based approaches to cognitive load in interpreting (Section 2.2.3) has already covered some studies that relate dependency distances and their association with disfluencies to cognitive processing issues. Several of these studies focus on the comprehension process, using dependency length measures in the source speech to predict the difficulty of the interpreting task (Jiang and Jiang, 2020). Focusing on the effect of long dependency distances during the comprehension process in interpreting, Jiang and Jiang (2020) link long dependency distances in the source text directly to an increase in cognitive load. In a sight interpreting task, with short and long dependency distance sentences they find that disfluencies occur significantly more frequently in the context of sentences with long maximum dependency distances, which are assumed to exceed short-term memory limitations, compared to short distances that fall within the short-term memory span. At the same time, the authors show an effect of shared syntactic structures of the two languages involved. In shared syntactic structures in source and target speech, exceedingly long dependency distances are assumed to be responsible for cognitive overload, whereas in non-shared structures, processing of long dependency distance relations and reformulating structures is the main cognitive challenge (Jiang and Jiang, 2020).

Liang et al. (2017) are the first to our knowledge to use the dependency distance approach as a tool investigating production processes in interpreting. They investigate dependency length differences across different interpreting types and compare dependency distances of consecutive and simultaneous interpreting to read-out translated speech (written-to-be-read production) in English from Mandarin as source language. Their findings show that the two interpreting modes have smaller dependency differences than the written-to-be-read production, with consecutive interpreting entailing the smallest dependency distances. The authors show that this difference is not source text triggered, nor can be linked to individual interpreter's style. Therefore, these results of shorter dependency distances in interpreting are assumed to be related to the significantly increased cognitive constraints in this production mode (Liang et al., 2017).

Other modes of multilingual communication, such as code-switching or written translation, show an inverse picture with dependency distances in these modes of communication that show longer dependency distances than in the monolingual setting (Fan and Jiang, 2019; Liang and Sang, 2022; Wang and Liu, 2013). Wang and Liu (2013) investigate code-switching with Chinese and English and compare dependency

distances in this setting to monolingual Chinese and English. Code-switching in this study refers to switching from one language to another within one clause or switching the language after one or several clauses (Wang and Liu, 2013). The multilingual setting generally leads to output with longer dependency distances than in monolingual production (Wang and Liu, 2013). Furthermore, the branching direction of the dependencies was shown to be affected by interference from the other language. Both findings are linked to higher processing costs in code-switching compared to monolingual communication (Wang and Liu, 2013). For translations, a source language effect in the target product is also assumed to lead to longer mean dependency distances. This was observed in two independent studies with written English translations based on Chinese sources compared to native English texts (Fan and Jiang, 2019; Liang and Sang, 2022). The authors link the effect of longer dependency distances in translations to source language influence, but also to a higher cognitive demand in producing a translation compared to monolingual production. However, the fact that Bizzoni et al. (2020) find dependency length minimisation for the translation direction German into English, but a reversed trend for English into German, also points to an effect of the translation process, resulting in different effects in the target product. German in general was found to be less optimal regarding DLM tendencies, which might influence the target translations in German. Furthermore, other known tendencies of translations such as explication, normalisation, etc. (Baker, 1996; Baroni and Bernardini, 2006; Volansky et al., 2015; Rubino et al., 2016) may influence dependency distances in translations. Looking at translations by L2 translators, the results show a simplification tendency (Liu and Afzaal, 2021). These translations with the source text in the native language being translated into a foreign language show less syntactic complexity, indicated by dependency measures (Liu and Afzaal, 2021). However, these effects may be rather influenced by L2 production.

Returning to spoken language use, Bizzoni et al. (2020) look not only at written translations, but also at English and German spoken originals and interpreting (based on German or English originals, respectively). Their dataset overlaps in parts with that used in this thesis. Bizzoni et al. (2020) look at summed dependency distance differences for each comparable dataset and observe shorter dependency distances in interpreting from English to German compared to German original speeches, but no differences between English interpreting compared to English spoken originals. This trend is opposite to the observed trend in the written dataset and is particularly interesting in the context of German, a language that is generally not as optimal as theoretically possible (Gildea and Temperley, 2010; Futrell et al., 2015; Dyer, 2023). German interpreting and, to a lesser extent, translations, too, show the most optimal dependency distances compared to original spoken or written German, respectively. One potential explanation could be the observed tendency of interpreters in German to imitate the clause word order of the source and produce the verb of a subordinate clause earlier than the expected verb-final position. This tendency was also observed by Collard et al. (2019) for subordinate clauses in German and Dutch interpreting. By

comparing mid-field length and extraposition in interpreting and original speech, the authors show that interpreters use extraposition significantly more frequently than original speakers as a means of reducing effort in the interpreting process (Collard et al., 2019). By producing the final verb earlier, overall dependency distance for the sentence is reduced compared to the non-extraposed variant. Generally, non-canonical word order choices may contribute to dependency minimisation trends (Yu et al., 2019).

Although English generally uses more optimal dependency distances (Gildea and Temperley, 2010; Futrell et al., 2015; Dyer, 2023), English interpreting from German as source language does not lead to less syntactic complexity compared to spoken English originals (Bizzoni et al., 2020). In simultaneous interpreting, sentence-by-sentence processing is assumed to have an impact on syntactic patterns used in the target speech (Liang et al., 2017), and therefore source language dependency length may have a much greater influence on the target speech than in the written context. Interpreting is seen as a linear process, and interpreters are generally assumed to avoid rearrangement of the word order of the source speech due to additional processing costs associated with rearrangement (Gile, 2017). Even in monolingual communication, dialogue participants are found to repeat syntactic structures previously heard in discourse, referred to as syntactic priming (Pickering and Branigan, 1999). Such an effect may even be larger in interpreting. In addition to the dataset used by Bizzoni et al. (2020), this thesis also investigates English interpreting from Spanish source speeches. It will be interesting to see if these dependency distance trends are a source language effect or rather an independent observation.

The last two studies to mention in the context of dependency measures in simultaneous interpreting are studies by Xu and Liu (2023) and Liu et al. (2023). The first study is very closely related to this thesis, investigating the simplification hypothesis in interpreting using dependency measures in simultaneous interpreting compared to English in non-mediated settings (original L1 and L2 speech) (Xu and Liu, 2023). Compared to this thesis, the study design differs mainly in the native language status of the interpreters. Interpreters in Xu and Liu (2023) interpret from their native language into an L2 language, while interpreters in EPIC-UdS (Przybyl et al., 2022b) work into their native language. Nevertheless, the main question on simplification in the interpreting product is addressed in a setup that compares L2 interpreting with L2 English production and English L1 production. The results by Xu and Liu (2023) on mean dependency distance are displayed in Figure 2.8, indicating that the shorter dependency distances in the English interpreting product produced by professional Cantonese native speaking interpreters are shorter than L2 speakers of English (Cantonese native). Compared to these two products of non native language use, native English production shows the longest mean dependency distances (Xu and Liu, 2023). Mean dependency distance in interpreting (L2I) differs significantly from native English production (L1O), while no significant differences are found for L2 originals vs. interpreting or L2 originals vs. native English production (Xu and Liu, 2023, p.7).

This finding is confirmed by findings from Liu et al. (2023) who, using a different set of syntactic complexity measures in a similar L2-interpreting setting, find reduced syntactic complexity in constrained language production and support the simplification hypothesis stating that interpreters reduce syntactic complexity due to the particular production conditions.

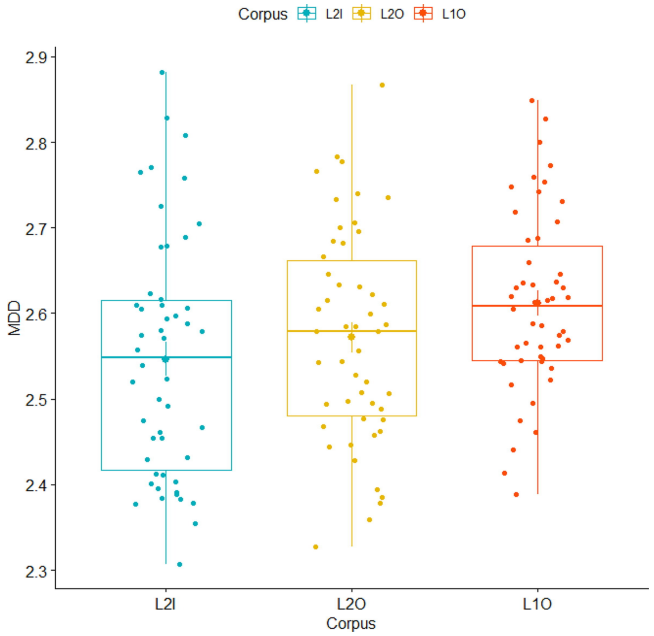


Figure 2.8: Mean dependency distance in interpreting (L2I) vs. non-native originals (L2O) and native English (L1O) (Xu and Liu, 2023, p.8).

To summarise the section on dependency length minimisation, it can be concluded that DLT is generally a suitable framework to apply for comprehension (Jiang and Jiang, 2020) as well as production processes (Liang et al., 2017; Bizzoni et al., 2020; Xu and Liu, 2023; Liu et al., 2023) in simultaneous interpreting. Although multilingual language production settings such as code-switching and translations have generally been found to lead to longer dependency distances (Wang and Liu, 2013; Fan and Jiang, 2019; Liang and Sang, 2022), the specific characteristics of simultaneous interpreting may require interpreters to encode their output efficiently. For syntactic complexity, this can be observed with shorter dependency distances in the interpreting product compared to monolingual production (Bizzoni et al., 2020; Xu and Liu, 2023; Liu et al., 2023). Generally, DLM can be related to syntactic choice (Temperley, 2007), however, in the cognitively particularly challenging task of simultaneous interpreting it is questionable as to what extent the interpreter is free to choose a more beneficial syntactical solution or is influenced by cognitive pressures and the linearity of source text items in production. Interpreter may be particularly forced to minimise dependency distances "for the sake of reducing memory burden" (Liu et al., 2017, p.1).

The languages involved in the interpreting process may also have an influence on potential dependency length minimisation trends in the interpreting product. For the languages investigated in this thesis, naturalistic language data in English tend to display more optimal dependency patterns than in German (Gildea and Temperley, 2007; Futrell et al., 2015; Dyer, 2023). Therefore, more dependency length minimisation tendencies in German interpreting compared to English interpreting (Bizzoni et al., 2020) are particularly interesting, and it remains to investigate to what extent the source language of the interpreting process influences the interpreting product.

2.4 Interpretese - known properties of interpreted language

This section aims to give an overview of the known properties of interpreted language, hereafter referred to *interpretese*. As this term was first used to describe properties of interpreting when compared to translations (Shlesinger, 2009), Section 2.4.1 takes this angle and shows the characteristics of interpreted language when compared to written translations. Sections 2.4.2 and 2.4.3 describe *interpretese* in the more narrow sense as used in this thesis, which is referring to linguistic properties of simultaneous interpreting when compared to original spoken language. A special focus is on specific properties that can be linked to simplification.

2.4.1 Properties of simultaneous interpreting when compared to written translations

Corpus-based work on written translation corpora goes back several decades (Gellerstam, 1986; Toury, 1991; Baker, 1993), however, what we know about interpreting from CBIS is still fairly limited. The potential of using interpreting corpora to compare source and target speech on a larger scale was first recognised by Shlesinger (1998) and soon repeated by the larger interpreting research community (Gile, 2004; Pöchhacker, 2004; Kajzer-Wietrzny, 2012; Bernardini et al., 2016). The time consuming nature of compiling corpora of spoken language and availability issues of large-scale interpreting data (Sandrelli et al., 2010) contribute to the fact that initially very few empirical interpreting resources became available. After the introduction of the first European Parliament Interpreting Corpus (EPIC, Bendazzoli and Sandrelli (2005)), various interpreting corpora have been published in recent years, mainly based on European Parliament proceedings (EPICG, Defrancq (2015); TIC, Kajzer-Wietrzny (2012); PINC, Chmiel et al. (2021); ESIC, Macháček et al. (2021)) or other political settings (SIREN, Dayter (2018); CEPIC, Pan (2019)). Interpreting corpora from other genres are still rare (e.g. HeiCIC, Kunz et al. (2021); NAIST, Doi et al. (2021); BST, Zhang et al. (2021)). Refer to Przybyl et al. (2022b) for a more detailed overview. The languages covered in these corpus resources are still very limited and

mostly include English as the source or target language. Therefore, what we know about the properties of interpreted language is still restricted to a limited number of language combinations and mostly to the political setting.

As corpus-based translation studies have been established for much longer, the first approach for analysing interpreting corpora was therefore taken from a translation context: studying translation-specific features as known from translation corpora research (*translationese*) in interpreted speech (Bendazzoli and Sandrelli, 2005; Monti et al., 2005; Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2012; Shlesinger and Ordan, 2012). The term *interpretese* as introduced by Shlesinger (2009) was also originally used to describe characteristics of interpreted language compared to translated language. The idea behind this approach of looking for similarities and differences between translation and interpreting was that by looking at interpreting data, translation scholars may be able to learn about the translation process and product just as interpreting scholars, by studying the slower and observable process and product of written translations, may be able to get insights about spoken translation under high-pressure circumstances (Pöchhacker, 2004). The current section gives an overview of the characteristic properties of simultaneous interpreting when compared to written translations.

Using the approach of investigating *translationese* features in interpreting and comparing written translation to interpreting reveals that, although there are similarities in the two modes of translation, interpreting is characterised by different lexical and grammatical tendencies (Shlesinger and Ordan, 2012; Karakanta et al., 2021; Bernardini et al., 2016; Ferraresi and Miličević, 2017). Many of these tendencies can be linked to *simplification* as a trend observed in translations which, when comparing translations to interpreting, can be observed stronger in interpreting than in translation. Lexical variety measured via TTR tends to be lower in interpreting than translations (Shlesinger, 2009; Shlesinger and Ordan, 2012; Bernardini et al., 2016). A lower level of lexical variety can generally also be linked to spoken language (Shlesinger, 2009; Shlesinger and Ordan, 2012), which is also true for many of the other features associated with *simplification*. Basic, simpler paradigmatic choices within the verb system in interpreting compared to translation can be observed, but this is not necessarily linked to lexical simplification, but can be associated with differences between spoken and written discourse (Shlesinger, 2009; Shlesinger and Ordan, 2012). Other features, such as the more frequent use of intensifiers and reduced forms (Przybyl et al., 2022a) or deictics (Przybyl et al., 2022a; Gast and Borges, 2023) in interpreting compared to translations, are also mainly indicators of the spoken mode. Furthermore, interpreting differs from translations at the lexical level in the choice of informal, colloquial lexical items (Shlesinger and Malkiel, 2005; Shlesinger and Ordan, 2012; Karakanta et al., 2021).

Studying the distributions of part-of-speech (POS) for interpreting compared to translations also reveals significant differences between interpreting and translation

(Shlesinger and Ordan, 2012; Przybyl et al., 2022a; Gast and Borges, 2023). This includes more frequent use of pronouns in interpreting, which are typically associated with orality (Shlesinger, 2009; Shlesinger and Ordan, 2012; Gast and Borges, 2023), adjectives that may rather be a source text shining-through feature (Shlesinger, 2009) or functions words (Shlesinger, 2009; Przybyl et al., 2022a) associated with explicitness or redundancy of spoken language such as prepositions, conjunctions, or copulas (Shlesinger, 2009). The more frequent use of conjunctions is also observed by He et al. (2016). There seems to be a difference in the type of conjunction used in the difference translation modes, with subordinating conjunctions being distinctive for translations which, in combination with relative pronouns and prepositions, points to more complex and longer syntactic structures in translations (Karakanta et al., 2021). Interpreting, on the other hand, is characterised by coordinating structures (Karakanta et al., 2021). Interpreters seem to have a preference to split up a sentence into several main clauses (Gast and Borges, 2023; He et al., 2016), pointing again to a more simplified discourse when comparing interpreting and translations. Related to this observation for conjunctions, mean sentence length in interpreting when compared to translation is found to be shorter (Bernardini et al., 2016).

On the other hand, interpreting is less characterised by possessives than translations (Shlesinger, 2009; Przybyl et al., 2022a). Distinct differences between the two modes of translation can also be found in the use of verbs and nouns, with interpreting using a more verbal style, while translations are characterised by nominal features (Shlesinger, 2009; Przybyl et al., 2022a; Gast and Borges, 2023). Gast and Borges (2023) find "a clear tendency for interpreters to "economize" on nouns" (Gast and Borges, 2023, p.11). The use of proper nouns is one of the major differences. Compared to written translation, simultaneous interpreting uses proper nouns very rarely and tends to use pronouns instead (Shlesinger and Ordan, 2012). However, this may depend on the languages and language combinations studied as He et al. (2016), who study translated and interpreted English based on Japanese sources, find pronouns less typical for interpreted English, while Shlesinger and Ordan (2012) study translated and interpreted Hebrew based on English sources.

Continuing on the lexical level, Ferraresi and Miličević (2017) observe that simultaneous interpreting uses significantly more infrequent modifier-noun combinations and fewer highly idiomatic modifiers of nouns than translations. The authors link these differences to differences in on-line and off-line language production. Regarding the choice of nouns used, interpreting is reported to use general nouns more frequently than translation (Bizzoni et al., 2021). Further differences at the lexical level are found by Bizzoni and Teich (2019) who study differences in lexical choice in translation and interpreting via neural semantic spaces (Word2Vec). Generally, they find looser word clusters for interpreting, but at the same time interpreting spaces displaying close nearest neighbours in the clusters. This points to high fidelity to source text items in the target with respect to unambiguous domain-specific words, whereas written translations show more diversity (more distant nearest neighbours

with more translation alternatives) in the same context (Bizzoni and Teich, 2019). More variation with regard to lexical choice in translation compared to interpreting is also confirmed by Przybyl et al. (2022a). This can be linked again to trends of *simplification* in interpreting compared to translation, which is also observed by traditional corpus linguistic measures for lexical variability (Shlesinger, 2009; Shlesinger and Ordan, 2012; Bernardini et al., 2016). Specifically, interpreted English is found to make more use of frequent words measured with traditional measures such as list heads (corpus-internal measure) and core vocabulary (measured compared to corpus-external database) (Bernardini et al., 2016), but also with a larger share of repeated content words (He et al., 2016).

Adverbs are found more frequently in simultaneous interpreting than in written translation (Gast and Borges, 2023), however, with a more limited range of adverbs used (Shlesinger and Ordan, 2012), lower variation pointing again to a more simplified discourse in interpreting. Differences in pragmatic use of adverbs are also indicated by the results of Bizzoni and Teich (2019) with no semantic equivalents for some adverbs found in the word clusters for interpreting.

Besides traces of *simplification* indicated by lower lexical variation in interpreting compared to translation observed with traditional corpus-based methods (Shlesinger, 2009; Shlesinger and Ordan, 2012; Bernardini et al., 2016) or information-theoretic measures on lexical choice (Bizzoni and Teich, 2019; Przybyl et al., 2022a), a greater tendency of simplification is also observed by Kunilovskaya et al. (2023b) who find lower information output in interpreting than translation for the same level of information input.

Simplification is not the only translationese tendency studied for interpreting. Other properties of translation that are found in interpreting include implicitation rather than explicitation in simultaneous interpreting when compared to translation with respect to the use of discourse connectives (Lapshinova-Koltunski et al., 2022, 2021), indicating a strategy that facilitates cognitive processing under severe time constraints. Greater use of transcoding as translation or interpreting strategy, such as the use of cognates (Shlesinger and Malkiel, 2005; Shlesinger and Ordan, 2012) and foreign words (Shlesinger and Ordan, 2012), can be linked to cognitive facilitation strategies resulting in the source text shining-through. Generalisation and omission seem to be particularly linked to interpreting, and restructuring leading to passivisation is particularly linked to specific language combinations such as interpreting from a head-final into a head-initial language (He et al., 2016).

Although many characteristics of interpreting when compared to translations can be linked to simplification trends, the parameters apply differently to different languages (Bernardini et al., 2016; He et al., 2016; Shlesinger and Ordan, 2012). In general, a stronger tendency of simplification is observed in interpreting than in translation (Bernardini et al., 2016). This may be due to the fact that simplification seems to be a feature of spoken discourse as well as meditated language use, and as interpreting is both, spoken and mediated, simplification can be observed

more strongly (Bernardini et al., 2016). Differences between the oral and written mode of translation are primarily differences in spoken and written language (Ferraresi and Miličević, 2017; Karakanta et al., 2021; Przybyl et al., 2022a; Gast and Borges, 2023). Therefore, Shlesinger and Ordan describe *interpretese* as being "more spoken than translated" (Shlesinger and Ordan, 2012, p.54). When aiming to find distinctive features of simultaneous interpreting and traces of cognitive processes, this thesis therefore only uses the spoken mode of translation and compares simultaneous interpreting to spoken language originals. *Interpretese* used in the following sections describes characteristic properties of interpreted language when compared within the spoken mode.

2.4.2 Interpretese from a lexical perspective

Unlike the studies on *interpretese* discussed above, the following section uses *interpretese* in the narrower sense, relating to specific properties of simultaneous interpreting only when compared to spoken originals, not to written translations. This comparison within the spoken mode allows to link indicators of lexical simplification to cognitive load during the production process (Plevoets and Defrancq, 2016). Previous research has shown that interpreted speech shows distinct lexical features compared to original speech. Attempts to study characteristics of interpreted language were initially mostly motivated by observations made for translations (cf. Section 2.1 for an overview of translationese features such as simplification, explicitation, shining-through, etc.). However, they use original spoken language as a basis for comparison rather than written translations (e.g. Sandrelli and Bendazzoli, 2005; Russo et al., 2012; Shlesinger, 2009; Kajzer-Wietrzny, 2012, 2015). This section focusses on *simplification*, but also covers other characteristics observed for simultaneous interpreting at the lexical level.

Lexical simplification was defined by Blum and Levenston as "the process and/or result of making do with less words" (Blum and Levenston, 1978, p.399). This mainly refers to less variation in the choice of words, but also to the information-carrying capacity of a text. Shallow statistical methods such as lexical density and lexical variation (Laviosa, 1998) can provide a first insight into differences between originals and interpreted texts from a lexical perspective. The features for simultaneous interpreting reported in the next section, studied in interpreting corpora and comparable original spoken data, are lexical density, lexical variation, and lexical sophistication, with the latter operationalised via the rate of word repetition, vocabulary size and depth (e.g. Kajzer-Wietrzny, 2012; Sandrelli and Bendazzoli, 2005; Dayter, 2018; Kajzer-Wietrzny and Ivaska, 2020; Ferraresi et al., 2018; Liu and Dou, 2023; Lv and Liang, 2019). This overview of studies is followed by further observations on lexis made within the context of *interpretese*.

Lexical density in simultaneous interpreting

Lexical density is the information-carrying capacity of a text or speech. It is defined as the number of lexical words divided by the number of running words in a corpus (Laviosa, 1998). The higher this number, the denser and therefore carrying more information a speech can be rated. Higher lexical density may lead to higher cognitive demands for information processing (Plevoets and Defrancq, 2018a). Lower lexical density can be seen as one indicator of simplification.

Most studies looking at lexical density in interpreting investigate interpreted English from different source languages and compare the results to original spoken English. Studies on EPIC (Sandrelli and Bendazzoli, 2005), covering the languages English, Italian, and Spanish, generally reveal only little variation in lexical density for original English compared to interpreted English from different source languages. In the comparable English and Italian subcorpora with interpreted English and interpreted Italian (from Italian or English and Spanish as source language), the authors note that lexical density is generally higher than in speeches originally delivered in the same language, however, with two exceptions to this trend of slight densification (Sandrelli and Bendazzoli, 2005; Russo et al., 2012). English interpreting from Italian sources is slightly denser than English original speech, but English interpreting from Spanish sources is less lexically dense than English original speech. For the comparable Italian subset, Italian interpreting with English as source is lexically denser than Italian original speech, while Italian interpreting with Spanish as the source language is less dense than Italian originals (Sandrelli and Bendazzoli, 2005). For EPTIC (Bernardini et al., 2016), a similar European Parliament corpus, covering the languages English, Italian, and French, the same pattern of no clear trend can be observed. Interpreted English with French as source shows significantly lower lexical density when compared to English originals, but interpreted English based on Italian sources is not significantly less dense than originals (Ferraresi et al., 2018). These results show that the language pairs involved in the interpreting process have an influence on the lexical density in the target speech. In a study only focusing on English based on European Parliament interpreting in the corpus TIC (Kajzer-Wietrzny, 2012), the lexical density in English original speeches is compared to English interpreted speeches based on French, Spanish, German and Dutch sources. In contrast to studies by Sandrelli and Bendazzoli (2005); Bernardini et al. (2016) and Ferraresi et al. (2018), interpreted speech for all language combinations covered in Kajzer-Wietrzny (2015) is found to be significantly denser, and therefore less simple and more informative than original spoken English. Furthermore, the trend of interpreted English being lexically denser than original spoken English is also confirmed for interpreted English based on Chinese sources (Lv and Liang, 2019; Xu and Li, 2022). Additionally, by investigating different genres of political speech, Xu and Li (2022) show an effect of the source speech genre on the lexical density in interpreting.

While some of these studies only report minor differences in lexical density, greater

differences in lexical density can be observed for the language pair English-Russian, investigated via the corpus SIREN (Dayter, 2018). Dayter (2018) reports higher lexical density in interpreted English based on Russian sources and a reversed trend for interpreted Russian based on English sources, showing that not only the language combination, but especially the interpreting direction influences the informativeness of the interpreting product.

A reversed trend to the results for English-Russian (Dayter, 2018) and most of the studies mentioned above (Bernardini et al., 2016; Kajzer-Wietrzny, 2015; Lv and Liang, 2019; Xu and Li, 2022) is found in a study based on interpreted English based on Russian sources in the United Nations Security Council compared to the original US English from the same setting (Liu and Dou, 2023). Liu and Dou (2023) not only measure the lexical density for each text but also compute lexical density values for different categories such as nouns, lexical verbs, adjectives, adverbs, numerals and pronouns, calculated as the ratio of the total number of, e.g. nouns in a text divided by the total number of tokens of that text. Their results show lower lexical density in interpreted English overall as well as for almost all subcategories investigated. With the exception of the lexical density of adjectives and adverbs, all investigated categories show lower lexical density and, with that, a higher degree of simplification in interpreted English compared to original US English (Liu and Dou, 2023). The corpus used by Liu and Dou (2023) is comparable to the data used in the other studies mentioned above in terms of register, and all corpora are situated in the political context. However, one major difference is the English variety spoken by the English speakers, with US English in the Liu and Dou (2023) dataset and Irish and British nationals in the European Parliament data (Sandrelli and Bendazzoli, 2005; Bernardini et al., 2016; Ferraresi et al., 2018; Kajzer-Wietrzny, 2015). It is questionable whether, in addition to language combination and text genre, also language variety may influence densification or simplification patterns.

Some of the different tendencies in lexical density observed for even the same language combination, interpreting direction, and text genre may also result from methodological differences in the study design which is generally comparing the share of function words to lexical words within a sentence, regardless of the overall length of the sentence. However, lexical density is assumed to be affected by sentence length (Xiao and Yue, 2010) and apart form Lv and Liang (2019) none of the studies mentioned above specifically state to consider sentence length in their design. With sentences generally shorter in simultaneous interpreting compared to original speeches in the same language (Łyda and Gumul, 2002; Bernardini et al., 2016; Gast and Borges, 2023), sentence length may be another factor generating bias in the results on lexical density in interpreting. Information density measured as *surprisal* may be better suitable to measure the information carrying capacity of a text, as, calculated via an n-gram approach, it is not sensitive to sentence length.

Furthermore, other properties of the corpora used in the studies above, in particular, in the mode of delivery of source speeches, may partially explain the different

results. For example, SIREN (Dayter, 2018) is compiled with a large share of impromptu delivered speeches, while EPIC (Sandrelli and Bendazzoli, 2005) and TIC (Kajzer-Wietrzny, 2012) rather use a mix of prepared, read-out speeches, and impromptu speeches.

Besides the interpreting process independent factors such as language combination or source speech properties that influence lexical density, particular interpreting production conditions may also have an effect on the information-carrying capacity of the target speech. Such factors may be "on-line" processing of the incoming source speech chunk by chunk, and imminent time constraints during simultaneous interpreting (Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2015). Time constraints in particular may be linked to avoidance of redundancies and the use of condensation techniques (cf. Section 2.2.2 for an overview of potential strategies used by interpreters), which can lead to higher lexical density in interpreting. More explicitation using lexical instead of referential cohesion (e.g. interpreters using lexical items instead of pronouns or auxiliary verbs) may be a further factor for more informationally dense speech in interpreting (Kajzer-Wietrzny, 2015). Another explanation may be interpreters using lexical cohesion instead of referential cohesion with a tendency to spell things out instead of using pro-forms or ellipsis (Shlesinger, 1995; Gumul, 2006). Such a tendency to explicitation would also increase the share of lexical words compared to function words and render interpreting more lexically dense.

To summarise the results on lexical density in simultaneous interpreting, it can be stated that no clear picture emerges. For interpreted English, there is no trend of simplification but rather densification (Kajzer-Wietrzny, 2015; Sandrelli and Bendazzoli, 2005; Ferraresi et al., 2018; Lv and Liang, 2019; Dayter, 2018). This includes the language combination studied in this thesis, interpreted English with German as source (Kajzer-Wietrzny, 2015). However, this trend is not confirmed across all source languages (Ferraresi et al., 2018) or language varieties (Liu and Dou, 2023). Simultaneous interpreting for other target languages shows a mixed (Sandrelli and Bendazzoli, 2005) or even reversed trend (Dayter, 2018). As lexical density, measured as the frequency of tokens that are not nouns, adjectives, adverbs, or verbs compared to all tokens, has been found not be a reliable predictor for texts being an original or a translation, but rather shows classification results close to chance level (Volansky et al., 2015), other measures that relate to the information-carrying capacity of texts such as *surprisal* may be more suitable to detect *interpretese* in general and *simplification* in particular.

Lexical variation in simultaneous interpreting

Simplification has traditionally also been investigated via the degree of lexical variation used in a text or speech. Low lexical variation is assumed to influence comprehension ease and can therefore be directly linked to the degree of simplification

of a speech (Kajzer-Wietrzny and Ivaska, 2020). The terms lexical variation, lexical variety, lexical variability, lexical diversity, or lexical flexibility are often used more or less interchangeably, sometimes even including lexical density, as discussed in the above section, or lexical sophistication referring to a speaker's command of less-commonly used words (Jarvis, 2013; Kajzer-Wietrzny and Ivaska, 2020). The term lexical variation used in this thesis refers to variation measured via the rate of word repetition, vocabulary size and depth (Kajzer-Wietrzny, 2012; Sandrelli and Bendazzoli, 2005; Dayter, 2018; Kajzer-Wietrzny and Ivaska, 2020). In interpreting studies, this has traditionally been operationalised via TTR, list heads and the proportion of high-frequency words compared to low frequency words, which indicate how lexically complex the speeches under study are (Kajzer-Wietrzny, 2012; Sandrelli and Bendazzoli, 2005; Dayter, 2018). High coverage by the list heads and high coverage of most frequent words is regarded as the most simplified. For spoken language in general, lexical diversity is generally lower in spontaneous speech (i.e. impromptu delivery) compared to prepared and read-out speech (Kajzer-Wietrzny and Ivaska, 2020).

Type-token ratio: Lexical variation can be measured via type-token ratio (TTR), with high scores relating to varied use of vocabulary, while low scores can be linked to a high level of lexical repetitiveness and a limited range of vocabulary used. Several studies find lower lexical diversity indicated by lower normalised TTR in interpreted speech compared to original spoken language: Liu and Dou (2023) observe more repetitive vocabulary for interpreted English based on Russian sources than for comparable English original speeches. The same trend is observed for interpreting from Cantonese to English with standardised type-token ratio (STTR) much lower in interpreting than original English speech (Lv and Liang, 2019), also observed across simultaneous interpreting of different genres of political speech events (Xu and Li, 2022). Ferraresi et al. (2018) compare TTR of English original speeches and English interpreting based on two different source languages. English originals are least repetitive with highest TTR while interpreting from French and Italian is less diverse. Interpreting based on French source show lowest TTR (Ferraresi et al., 2018). However, these observed differences are not statistically significant (Ferraresi et al., 2018). Bizzoni et al. (2021) confirm this general trend focusing on nouns in interpreted English based on German compared to original English. Summarising the results on TTR, a general trend of interpreting being more repetitive, using a narrower range of vocabulary than original speeches can be observed.

List heads: Another way to investigate repetitiveness of a dataset is by comparing the 100 most frequent words with the total amount of words in the dataset. If the 100 most frequent words are used very frequently overall in the dataset, lexical variety is regarded low. The same words are then used over and over again. This corpus-internal measure, referred to as *list heads*, is an indication of repetitive use of lexis - higher values indicate more repetitiveness and higher lexical simplification

(Laviosa, 1998).

List head studied in various interpreting corpora show very mixed results, especially for original English compared to interpreted English from different source languages. In EPIC, interpreted English, regardless of the source language (i.e. Italian and Spanish) uses high-frequency words more frequently compared to original spoken English, indicating less lexical variety in interpreting (Sandrelli and Bendazzoli, 2005; Bernardini et al., 2016). This is also true for English interpreting based on French and Italian, showing significantly higher list head coverage than English originals (Ferraresi et al., 2018). The same results for interpreted English based on Spanish sources are also observed in the TIC corpus, showing greater repetitiveness compared to English originals (Kajzer-Wietrzny, 2015). Interpreted English with French as source, on the other hand, shows the opposite trend of being less repetitive, whereas no significant differences in list head coverage are observed when comparing English originals to interpreted English based on German or Dutch sources (Kajzer-Wietrzny, 2015). Less repetitive lexis is also found in English interpreting based on Russian source speeches compared to original spoken English in the SIREN corpus (Dayter, 2018), as well as for interpreted English based on Cantonese originals when compared to original English speech (Xu and Li, 2022; Lv and Liang, 2019).

Looking at other target languages and language combinations, the results also show a mixed picture. For interpreted Italian based on English or Spanish sources, more lexical variety is observed using the measure of list heads than for original Italian speech (Sandrelli and Bendazzoli, 2005). This trend is reversed, again, for interpreted Russian based on English sources with interpreted Russian using less varied lexis than original Russian speech (Dayter, 2018).

These different and sometimes opposite trends are difficult to interpret, with some authors partly linking the effect with small data size (Sandrelli and Bendazzoli, 2005) or language effects (Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2015). However, the opposing tendencies for English interpreting based on different source languages (Sandrelli and Bendazzoli, 2005; Bernardini et al., 2016; Kajzer-Wietrzny, 2015; Dayter, 2018) rather point to a shining-through effect, linked to the language combinations involved in the interpreting process. For example, some language pairs and interpretation directions might be more affected by a tendency to use cognates such as English and French being typologically very close. Therefore, more lexical borrowing may be used when interpreting between these two languages (Kajzer-Wietrzny, 2015; Shlesinger and Ordan, 2012). Greater use of cognates and more unusual lexical formulations are assumed to increase lexical variety, making the language more diverse and thus less simple (Kajzer-Wietrzny, 2012, 2015). Kajzer-Wietrzny (2012) also suggests that the mode of delivery of a source speech impacts the degree of repetitiveness of the interpreted speeches, with interpretations of impromptu speeches generally showing a higher degree of repetitiveness.

Overall, no clear trend is observed via list heads as an indicator of repetitive use of the language. Interpreted speech does not seem to be more repetitive or lexically less

varied than in original speech. This feature may be rather dependent on language combinations.

Core vocabulary coverage: A similar approach to measuring the range of vocabulary used in a corpus is to measure the percentage of text covered by the most frequent words in a language taken from a reference corpus and compare the result to the vocabulary used in the corpus. This corpus-external measure is also referred to as *core vocabulary coverage* (Bernardini et al., 2016).

Higher core vocabulary coverage can be observed in interpreted English compared to originally spoken English (Bernardini et al., 2016). However, this trend, again, depends on the source languages interpreted English is based on. English interpreting based on French sources show higher core vocabulary coverage than English originals (Ferraresi et al., 2018). Although such a trend is also observed for interpreted English based on Italian, this trend is not significant (Ferraresi et al., 2018). Interpreted English based on Chinese originals when compared to original English speeches uses more diverse lexis, as indicated by this corpus-external measure (Lv and Liang, 2019).

Lexical sophistication: Other measures linked to lexical variation relate to the degree of lexical sophistication, such as mean word length or the use of advanced vocabulary like rare or infrequent words. High lexical sophistication is believed to be linked to strong cognitive abilities in handling the cognitive load (Liu and Dou, 2023).

Looking at the mean word length as the most simple approach to lexical sophistication, it can be observed that interpreting is characterised by using shorter words than original speech (Karakanta et al., 2021). Investigating the use of advanced vocabulary in simultaneous interpreting, Liu and Dou (2023) approximate lexical sophistication by the proportion of words in a dataset that are part of the Academic Word List (Coxhead, 2000) and the proportion of off-list words. Both indices show less lexical sophistication in interpreted English based on Russian compared to original English speech (Liu and Dou, 2023). Frequency of the type of words used in interpreting is also considered in a study by Bizzoni et al. (2021). The authors show that interpreters use general words, approximated via a shorter or longer path in the WordNet hierarchy (Fellbaum, 2010), more frequently than original speakers. This may be linked to a general frequency effect, but also to retrieval advantages for general words. In line with these trends of interpreting using less sophisticated vocabulary, Lv and Liang (2019) observe that hapax legomena, word forms that appear only once in the corpus, occur less frequently in simultaneous interpreting than original English.

Generally, a trend of lower lexical sophistication in simultaneous interpreting can be observed.

Formulaicity: Beyond the unigram approaches to lexical variation mentioned above, several studies use frequency of n-grams to investigate variation in interpreted

language. In an exploratory corpus approach that looks at n-grams with 5 words and more, Aston (2018) observes that interpreters in the European Parliament rely heavily on recurrent formulaic phraseologies. This is confirmed by Kajzer-Wietrzny and Grabowski (2021) who find that interpreting uses a higher share of most frequent fixed or semi-fixed multi-word units than original spoken language. Spoken original language produced by native speakers uses fewer frequent distinct bigram types compared to non-native spoken production, and interpreters working into their non-native tongue produce the highest amount of recurrent multi-word units. This may be linked to the double cognitive constraint of two types of constraint communication: interpreting and speaking in a foreign language (Kajzer-Wietrzny and Grabowski, 2021).

Generally, interpreters experience higher cognitive load when faced with difficult source speech properties, such as a higher degree of preparedness or a fast source speech delivery type. In such difficult conditions, interpreters use more formulaic language (Kajzer-Wietrzny and Grabowski, 2021). A higher rate of delivery contributes to more, but not significantly more, most frequent bigram types in the interpreting output. Furthermore, speeches with a lower degree of planing such as impromptu delivered original speeches and interpreted language show more formulaicity than prepared and read-out speeches (Kajzer-Wietrzny and Grabowski, 2021).

Dayter (2019) also observes a tendency of interpreters to choose a formulaic solution in cases where the source speech uses an unusual ad-hoc description. Lexical bundles (trigrams) that are very conventional for the target language occur statistically more frequently in interpreted English than in original English (Kajzer-Wietrzny, 2012). However, a source language effect is also found regarding the use of reoccurring linguistic patterns (Kajzer-Wietrzny, 2012). Both authors link this particular use of formulaic expressions to normalisation tendencies in interpreting (Dayter, 2019; Kajzer-Wietrzny, 2012). Nevertheless, it is listed here among simplification tendencies in interpreting, as the occurrence of a set of fixed trigrams in the target speech makes interpreting more repetitive and expected, and therefore should ease processing by the audience of the interpretation.

To summarise studies on lexical variation and sophistication, findings point towards a simplification tendency in simultaneous interpreting regarding these measures. Studies generally find interpreting to be more repetitive and less diverse, using a narrower range of vocabulary than original spoken language (Russo et al., 2012; Kajzer-Wietrzny and Ivaska, 2020; Xu and Li, 2022). Some authors interpret their results as "diverging patterns" for lexical simplification (Ferraresi et al., 2018), others see the simplification hypothesis generally confirmed, however, via different features. English interpreting used more frequent words and text-internal repetitions, while Italian uses syntactic simplification and a greater share of function words (Bernardini et al., 2016).

Some of the measures used to investigate lexical variation give a clearer picture

than others. While TTR, measures on lexical sophistication and studies on formulaicity point toward a simplification trend with more repetitive lexis in simultaneous interpreting (Ferraresi et al., 2018; Lv and Liang, 2019; Bizzoni et al., 2021; Kajzer-Wietrzny and Grabowski, 2021; Xu and Liu, 2023; Liu and Dou, 2023), the measure of list heads or core vocabulary coverage gives rather mixed results (Kajzer-Wietrzny, 2012, 2015; Dayter, 2018; Ferraresi et al., 2018; Lv and Liang, 2019; Xu and Liu, 2023) with the proportion of high-frequency words in simultaneous interpreting using less simplified lexis (Kajzer-Wietrzny, 2012).

This thesis uses the information-theoretic measure of *surprisal* to investigate lexical simplification as it combines the measures traditionally used for this purpose. It gives the information density of a token that represents the information-carrying capacity of a text, traditionally measured as lexical density. Furthermore, it considers frequency effects and formulaicity by considering frequency of a token in an n-gram context.

Using a measure of average segment surprisal, Kunilovskaya et al. (2023b) show that information density in interpreters' output is conditioned by information density of the source. Moreover, the authors link "significantly lower information density in spoken mediated production compared to non-mediated speech in the same language" (Kunilovskaya et al., 2023b, p.608) to a possible simplification effect in interpreting. While no effect on information density is found for read-out speeches compared to impromptu speeches, source speech delivery rate affects the informativity of the target. The dataset used in this study is identical to the one used in this thesis; however, instead of using average segment surprisal, this thesis compares average surprisal of distinct POS to investigate a potential simplification effect on lexical level.

Further lexical translationese properties in simultaneous interpreting

Although *simplification* is the translation universal investigated most for interpreting, other translationese tendencies are also studied as potential characteristics for *interpretese*.

As we have seen in the studies on *simplification* in the previous section, source language and text properties play a vital role in the properties of interpreters' output. Traces of source speech *shining-through* in interpreters' output is observed by Dayter (2018), investigating POS distributions for interpreted English and Russian. In particular, the distributions for nouns and verbs point to a clear influence of the source language. Interpreted English based on Russian sources is observed to be more nominal and influenced by the nominal typology of Russian. On the other hand, interpreted Russian based on English sources is more verbal, which can be interpreted as a *shining-through* effect of the verbal influence of English (Dayter, 2018). In a further study, Dayter (2019) looks at collocations and finds further traces of *shining-through* in simultaneous interpreting for certain verb-adverb patterns. The author finds that interpreters stay close to the source language input at the expense of correctness and idiomaticity (Dayter, 2019). Generally, *shining-trough* tendencies can make the tar-

get speech less simple for a target audience, as target speech with source language traces would be more unexpected and, therefore, more difficult to process compared to original speech in the same language.

Some *shining-through* effects of the source language may also be linked to another traditional translation universal. The use of nominal reference instead of pronominal reference, known as *explicitation* in the literature (Lapshinova-Koltunski, 2015), may also be linked to the distribution of corresponding POS in the source speeches with opposing trends in different language combinations. Interpreted English based on Russian sources is more explicit as it uses nominal instead of pronominal reference, while the opposite is true for interpreted Russian based on English sources (Dayter, 2018). For the language combination English-Hebrew, a lack of *explicitation* and rather traces of *implication* are observed (Shlesinger (1989) cited in (Pym, 2007)). A preference for pronominal reference rather than explicitly referring to the noun is also observed for the language combination German-Spanish (Karakanta et al., 2021). Such *implicitation* may be linked to time constraints as long words (nouns are often longer) are replaced with short pronouns (Karakanta et al., 2021).

Kajzer-Wietrzny (2018) finds increased *explicitation* by interpreters adding an optional connective *that* after reporting verbs (compared to non-native originals). Potential causes of this type of explicitation could be, among others, the higher cognitive complexity of simultaneous interpreting as explicit grammatical solutions are found to be more frequent in cognitively complex environments (Olohan and Baker, 2000) or linked to *normalisation* as a tendency of interpreters to be overcorrect as well as source language patterns reflected in the target (*shining-through*) (Kajzer-Wietrzny, 2012). *Explicitation* in interpreting is also observed via a tendency towards increased cohesion with the use of more cohesive devises (Kajzer-Wietrzny, 2021; Defrancq et al., 2015). Such shifts in cohesion in interpreted speech, be it omission that may be linked to *implicitation* or repetitions, linked to *explicitation*, provide a direct link to *simplification* when interpreters use complete lexical reiteration or repetition, using superordinate terms but almost never use synonyms (Łyda and Gumul, 2002).

Investigating other means of explicitness, such as linking adverbials or connectives, no trend of *explicitation*, but rather *implicitation* in simultaneous interpreting is observed by omitting cohesive ties (Lapshinova-Koltunski et al., 2022; Kajzer-Wietrzny, 2012). These effects on the lexical level may be linked to "the linearity constraint" (Shlesinger, 1995) with the speech being unveiled to the interpreter only fragment by fragment, One needs to be very cautious not to misinterpret the speaker's intention (Kajzer-Wietrzny, 2012).

Further *explicitation* tendencies observed in the literature are ellipsis and substitution replaced with some form of lexical cohesion (Shlesinger, 1995). Pym (2007) suggests this may be linked to a risk-reducing strategy as interpreters know less on the subject matter than the audience and the speaker. Leaving their rendition more implicit reduces the risk of a wrong translation. These *explicitation* shifts are found

to be mostly subconscious additions, repetitions, and other cohesive shifts (Gumul, 2006).

Explicitation tendencies in general can be assumed to make the interpreting product easier to process by the audience as cohesive links are spelt out rather than having to be inferred by the audience.

The idea of a further translation universal, adapting the target speech to target-culture norms, is rejected by Shlesinger (1989) cited in (Pym, 2007)). Rather than *normalisation*, Shlesinger (1989) proposes a *equalising universal* for interpreting where orality of markedly oral texts and literateness of markedly literate texts are reduced during the simultaneous interpreting process. *Equalisation*, sometimes also called *levelling out*, can explain certain properties of simultaneous interpreting, considering that some source speeches contain more literate, others more oral characteristics (Dayter, 2019). Interpreters are found to complete unfinished sentences, give a grammatical rendition of an ungrammatical source utterance, and omit false starts and self-corrections of the source (Schlesinger, 1991). Investigating recurrent word combinations or collocations via certain POS-chains, Dayter (2019) also finds support for the *equalising universal* as certain POS patterns are more linked to spoken language, others can be linked to more literate language, and both tend to move more towards the centre of the continuum. In general, the range of the oral-literate continuum is reduced in simultaneous interpreting (Dayter, 2019). However, Shlesinger's (1989) results do not fully support the equalising hypothesis (Pym, 2007). Literate texts were confirmed to become more oral when interpreted. However, orally marked texts do not always move towards the centre of the oral-literal continuum, but even show some characteristics of becoming even more orally marked than the originals they are based on (Pym, 2007). Overall, interpreting products from different source languages are more similar to each other than compared to original speeches in the same language (Kajzer-Wietrzny, 2015).

Summary on lexical simplification patterns in simultaneous interpreting

Previous studies on lexical simplification show mixed tendencies on the different measures investigated for simultaneous interpreting. No simplification trend, but rather densification, can be observed for the measure of lexical density (Kajzer-Wietrzny, 2015; Sandrelli and Bendazzoli, 2005; Ferraresi et al., 2018; Dayter, 2018; Lv and Liang, 2019). Interpreter's output is conditioned by source text input and it is reassuring to see that time pressure and other constraints of the interpreting process do not lead to a reduction of lexical items, but rather fewer function words that carry less information. As for simplification measures related to lexical variation, a trend of more repetitive and less diverse lexis, a narrower range of vocabulary, and lower sophistication in simultaneous interpreting than original spoken language can be observed (Russo et al., 2012; Kajzer-Wietrzny and Ivaska, 2020; Xu and Li, 2022),

however, with some exceptions to this trend (Ferraresi et al., 2018; Dayter, 2018, among others).

These mixed results show that simplification is not a simple phenomenon, but rather an interplay of different factors such as language pair, working direction or source speech properties like text genre, lexical density in the source, delivery rate, or mode of delivery. While some of these factors can be linked to the languages involved in the interpreting process and textual factors of the source speech, others may rather be the result of the interpreting process as such. Assumed simplification tendencies linked to cognitive optimisation pressures may be counterbalanced by interpreting process-external factors.

Reversed tendencies for the same language pair but reverse interpreting direction, as observed in Dayter (2018), clearly show the influence of source language properties on the interpreting product. Properties of the source speech, such as text genre, lexical density of the source speech, delivery rate, or mode of delivery, may also influence simplification tendencies in the target. Mode of delivery of a source speech has an influence on repetitiveness (Kajzer-Wietrzny, 2012). In particular, impromptu delivered original speeches are found to be less diverse than prepared and read-out speeches, containing fewer rare words, fewer content words, and more synonyms (Kajzer-Wietrzny and Ivaska, 2020). Interpreting from read-out speeches would be expected to lead to a more diverse range of vocabulary used; however, the opposite can be observed Kajzer-Wietrzny (2015). Furthermore, source speech delivery rate influences simplification tendencies in the target (Kunilovskaya et al., 2023b). While these results can potentially be linked to an *equalising effect* (Shlesinger, 1989), Kajzer-Wietrzny (2015) links these results to interpreters lexicalising cohesive links, expressed as pronouns in the original speech but as lexical forms in interpreting. This would increase the range of vocabulary and would be seen as a less simplified output. However, at the same time, an explicitation effect can be observed where a cohesive link is expressed more explicitly in interpreting than in the source. This type of explicitation may facilitate processing for the listener.

Other factors influencing the simplification tendencies in interpreting can be directly associated with the interpreting process as such. Lexical simplification can be linked to cognitive load with interpreters simplifying vocabulary when their cognitive capacity is close to saturation (Lv and Liang, 2019). Cognitive pressure may lead to the use of simpler vocabulary, including commonly used and stereotypical words and fixed or semi-fixed multiword units (Liu and Dou, 2023).

Working direction – interpreters working from A to B or from B to A – may also influence simplification due to different cognitive demands working into or from ones native language (Xu and Li, 2022). Mixed findings for different interpreting directions (A to B and B to A) may be the result of over-corrections when working into a B-language, which may not only increase lexical density but also lead to more varied vocabulary used (Dayter, 2018). Other authors assume working direction from A to B to create higher cognitive load and therefore a higher degree of simplification (Xu

and Li, 2022).

Generally, lexical simplification may be linked to professional interpreter's (conscious) ability to apply certain strategies, prioritise multiple cognitive demands and manage cognitive load (Liu et al., 2004). Interpreter's adherence to linear constraints (Shlesinger, 1995) may also result in certain patterns in interpreting such as a higher proportion of function words leading to lower lexical density (Liu and Dou, 2023).

To conclude, simplification pressures due to constraints of the interpreting process may be counteracted by interpreting process-external factors. Descriptive corpus-based approaches to lexical simplification give mixed results. While unigram-approaches to lexical simplification show diverging patterns, studies on formulaicity considering n-grams are a first approach to look beyond single words. These results show a clearer picture of repetitive patterns in interpreting (Kajzer-Wietrzny, 2012; Aston, 2018; Kajzer-Wietrzny and Grabowski, 2021; Dayter, 2019). This thesis aims to investigate lexical simplification via the information-theoretic measure of surprisal as, besides relating to the information-carrying capacity of a text, it considers the contexts in which a word occurs and can be seen as measuring lexical variation in context. Moreover, it provides a direct link to cognition as it quantifies cognitive load as an indicator of the effort required to comprehend or produce language (Hale, 2001; Teich et al., 2020).

Most of what we know about *interpretese* relating to lexical properties is based on corpus-based studies with English as source or target (Sandrelli and Bendazzoli, 2005; Russo et al., 2012; Ferraresi et al., 2018; Dayter, 2018; Lv and Liang, 2019; Xu and Li, 2022; Liu and Dou, 2023, amongst others). Only very few studies include German as a source language (Kajzer-Wietrzny, 2012, 2015; Lapshinova-Koltunski et al., 2022; Kunilovskaya et al., 2023b) and even fewer German as a target language (Lapshinova-Koltunski et al., 2022; Kunilovskaya et al., 2023b). This thesis aims to fill this gap by specifically investigating *interpretese* in general, and specifically *simplification* for German-English and English-German.

2.4.3 Interpretese from a syntactic perspective

While numerous studied have investigated lexical simplification patterns in simultaneous interpreting, studies investigating simplification at the syntactic level are much rarer (Liu et al., 2023). Syntactic measures can provide deeper insight into the interpreting process (Liu et al., 2023) and can also provide a direct link to cognitive resources management during simultaneous interpreting (Seeber and Kerzel, 2012b; Hild, 2011). Syntactic complexity is relevant for the interpreting process regarding the source speech as well as the target speech. High syntactic complexity in the source, which cannot be influenced by the interpreter, may lead to increased cognitive load (Jiang and Jiang, 2020; Chmiel et al., 2024). On the other hand, syntactic complexity of the target speech can be modulated by the interpreter in order to relieve cognitive

burden (Chmiel et al., 2024). In particular, the application of certain production strategies such as chunking may be beneficial to reduce syntactic complexity and cognitive burden (Ma et al., 2020). Furthermore, it is believed to be beneficial for interpreters' cognitive resources management when source and target speech share morphosyntacic symmetry (Seeber and Kerzel, 2012b). Producing complex sentences in the target may (Chmiel et al., 2024) or may not (Plevoets and Defrancq, 2020) lead to an increase in interpreter's cognitive load, depending on the operationalisation of cognitive load.

Sentence length and syntactic complexity in simultaneous interpreting

A first indicator of syntactic simplification can be the length of the translation unit, operationalised by sentence length, in simultaneous interpreting, assuming that longer units are generally considered more complex than shorter ones (Liu et al., 2023). In EPIC, interpreting is generally found shorter on the word level than the source speeches interpreting is based on, including English, Italian, and Spanish source speeches and interpretations into all possible combinations and directions[2] (Russo et al., 2012). The only exception to this trend of shorter sentences in interpreting was found for interpreting from Italian into Spanish, which the authors explain via increased use of function words for this language pair. Besides this parallel corpus-based approach, (Bernardini et al., 2016) also observe significantly shorter sentences in comparable interpreting and spoken originals of the same language. The same trend is also observed for the language pair English to German, with Gast and Borges (2023) observing interpreters splitting complex clauses in the source into two independent main clauses in the target.

These results show that sentence length in interpreting can be directly linked to the level of embeddedness, a further syntactic measure of complexity, including frequency and depth of embedding. Some structures, such as coordinated phrases and complex nominals, are considered more complex than others and the frequent occurrence of such structures is linked to higher syntactic complexity (Lu, 2010). Karakanta et al. (2021) find original speeches in Spanish to be characterised by nominal structures and complex main clauses, while relative pronouns are found to be characteristic for interpreting, pointing towards subordination in interpreting (Karakanta et al., 2021). While syntactic embeddedness, such as subordination, can be cognitively more demanding, interpreters tend to encode such structures explicitly by using optional complementizer *that* more than original speakers (Kajzer-Wietrzny, 2012). Such syntactic explicitation can be linked to syntactic simplification of the target speech as it may help the audience to process interpreters' output more easily. Other connectives, especially those that do not contribute to the propositional content or can easily be recovered, are sometimes omitted in interpreting, resulting in two clauses merging

[2](interpreting from English into Italian and Spanish; from Italian into English and Spanish; and from Spanish into Italian and English)

into one (Łyda and Gumul, 2002).

A more sophisticated approach to measure syntactic complexity in simultaneous interpreting vs. original speech is taken by Liu et al. (2023). The authors use 14 different syntactic measures, such as the length of the production unit, the amount of subordination and coordination, phrasal sophistication and overall sentence complexity, using measures of the L2SCA tool (Lu, 2009). They find simultaneous interpreting (from interpreter's native language into non-native language) to be syntactically more simple than original English production by native speakers. For 12 of the 14 measures, the differences were found to be significant. Only two measures, mean length of clause and number of coordinate phrases per clause, did not yield significant results (Liu et al., 2023). As their study design compares original native production to simultaneous interpreting into a non-native language, their dataset also includes original non-native production data. This allows for verifying that results found for simultaneous interpreting are not only due to the non-native production mode. Particularly, their results on the subordination measures clearly state that simultaneous interpreting is least complex in terms of subordination use, compared to original native and non-native production. Also, regarding the measures on phrasal and overall sentence complexity, lower results on these complexity measures can clearly be linked to simultaneous interpreting, and not only to non-native language production (Liu et al., 2023). Their results on coordination, on the other hand, can be linked to non-native language use rather than being a result of interpreting. When interpreters produce longer sentences, this is primarily due to coordination at the phrase level rather than coordination at the sentence level (Liu et al., 2023).This use of phrasal coordination can improve clarity and explicitness of the target speech (Liu et al., 2023). Overall, Liu et al. (2023) find clear evidence the interpreting is syntactically simpler than native and non-native originals.

The results of interpreting generally producing shorter and less complex syntactic units, as cited above, can clearly be linked to the interpreting process, in line with the interpreting strategy of *chunking* (cf. Section 2.2.2), resulting in the interpreting product being syntactically more simple than original production. However, as observed for lexical measures in Section 2.4.2, source speech influence, shining-through of source speech patterns, can also be observed in the interpreting product.

Alonso-Bacigalupe (2010) observe that interpreters working between English and Spanish often produce syntactic structures in the target that are parallel to the source text structures. Interpreters are also found to maintain the question structure of the original for their rendition in the target (Sandrelli, 2018). Information seeking questions (wh-questions and modal polar questions) and confirmation seeking questions (yes/no questions, choice questions, declaratives, and imperatives) are mostly rendered using the same question type in the interpretation (Sandrelli, 2018).

Bizzoni et al. (2020) also investigate shining-though, but focus on the language

pair German-English. The authors compute the language model perplexity for POS sequences of different language models on unseen data from different categories to assess structural differences or similarities between German spoken originals, interpreted German based on English sources and English spoken originals. The approach can be used to find tendencies of shining-through (if interpreted German is similar to English originals) or normalisation (if interpreted German is similar to German originals) in interpreting. However, the authors find German interpreting sequences to be perplexing for all language models, also compared to the spoken German model. This shows that POS sequences in interpreting are very different from spoken originals and not necessarily only influenced by source language patterns. The authors hypothesise that this could be due to structural over-simplification, as a possible effect of the cognitive pressure on interpreters.

Dependency locality measures in simultaneous interpreting

So far, few studies have investigated *interpretese*, defined as specific properties of simultaneous interpreting when compared to original spoken language, via the approach of Dependency Locality Theory (DLT) (Gibson, 1998, 2000). This approach using the linear distance between two syntactically related words in a sentence can be seen as index of syntactic complexity and, at the same time, reflects cognitive constraints during different tasks (Liang et al., 2017).

Using the approach of DLT (cf. Section 2.3.2), Bizzoni et al. (2020) investigate average summed dependency distances per sentence length for English spoken originals and English interpreting based on German, as well as German spoken originals and German interpreting based on English. The authors find German interpreting based on English to be less complex than German originals, but there are no differences for English interpreting based on German compared to English originals. This result is interpreted as shining-through of the source language the interpreting product was based on (Bizzoni et al., 2020).

Also using the DLT approach, Xu and Liu (2023) investigate syntactic simplification and its cognitive implications in simultaneous interpreting using mean dependency distance and dependency direction. Unlike the setup used in this thesis, which is comparing simultaneous interpreting from a non-native language into the native language, Xu and Liu (2023) study simultaneous interpreting with the source speech in interpreter's native language into interpreter's non-native language. Their setup includes native English original speeches, non-native English original speeches (produced by native Chinese speakers), and English interpreted speeches produced by native Chinese speakers. Their results show that interpreting uses the lowest mean dependency distance, followed by non-native original production with native original speech using significantly longer dependency distances (Xu and Liu, 2023). Regarding the dependency direction, a source language effect is found, with dependency direction in non-native English and interpreted English being influenced by Chinese word

order patterns (Xu and Liu, 2023).

Support for the simplification hypothesis regarding syntactic complexity measures via mean dependency distance is also given by Chmiel et al. (2024) in a parallel corpus-based approach. The authors find lower syntactic complexity in interpreted speeches than in the source speeches on which they were based. These results for the language pair Polish-English were observed for both interpreting directions, although the source speeches in English were found to have longer dependency distances than Polish source speeches (Chmiel et al., 2024). They also show that speech planning has an effect on syntactic complexity with impromptu speeches using longer dependency distances than prepared and read-out source speeches. However, this has no effect of simplification tendencies in interpreting with interpreters simplifying syntax, regardless of delivery mode (Chmiel et al., 2024). Both studies, Xu and Liu (2023) and Chmiel et al. (2024), do not mention taking into account differences in sentence length in their study design. As previous research has shown that interpreters produce shorter sentences than original speakers (Bernardini et al., 2016; Liu et al., 2023), this may have implications on the measure of dependency distance.

A cognitive approach to studying the interpreting product is also taken by Collard et al. (2019). The authors do not refer to the dependency length minimisation approach (Gibson, 2000) explicitly, but their approach also considers long syntactic relations to be cognitively more challenging as open relations have to be maintained in memory until they are closed. Their study design, looking at the verbal brace in subordinate clauses in German and Dutch, and investigating a potential tendency of interpreters producing the final verb in an SOV structure earlier and extrapose certain elements after the verb, can be directly linked to dependency locality (Collard et al., 2019). The study compares German and Dutch originals to German and Dutch interpreting based on French originals, which is interpreting from an language with SVO pattern to a language that uses SOV in subordinate clauses as default (Gerritsen and Stein, 1992; Eisenberg, 1994).

(14) ... *die* vor allem *zu tun haben* mit der Einrichtung einer europäischen Aufsichtsbehörde / oder auch mit / der europäischen Registrierung und Zulassung euh euh von euh Wertpapierverwaltungsgesellschaften
(EN: ...which mainly *have to do* with establishing a European supervisory body / or also with European registration and certification of portfolio management companies

Example 14 (Collard et al., 2019, p.704), shows a German interpreted subordinate based on a French original, in which the distance between the verbal group of the subordinate *zu tun haben* to the relative pronoun it depends on (*die*) is kept short with only two words in the midfield. The interpreter extraposes 16 words that could theoretically have been included in the midfield. The authors show that in both languages studied interpreters produce significantly shorter midfields and therefore

shorter dependency distances between the verbal group in the subordinate and the head it depends on compared to original speakers in the respective languages (German and Dutch) (Collard et al., 2019). Instead, interpreters tend to extrapose clausal elements that could theoretically have been placed in the midfield after the verbal group. They do so more frequently than original speakers in the same language, although this trend is only significant for interpreted German, and not for Dutch.

This result of shorter midfields can directly be linked to the associated cognitive load which is required to retain information to produce the verbal brace at the end of the sentence. Collard et al. (2019) also look into potential triggers of their findings and observe a source language influence as a probable cause. Interpreting SVO into SOV with maximised extrapostions is structurally more similar than producing SOV without extraposition, and indeed, around 90% of all extrapositions in their study were triggered by structurally similar clause word order of the source (Collard et al., 2019). The authors interpret these traces of shining-through as interpreters prioritising memory management over target acceptability.

To summarise the properties of syntactic *interpretese*, we can see evidence of syntactic simplification (Russo et al., 2012; Bernardini et al., 2016; Gast and Borges, 2023; Liu et al., 2023; Bizzoni et al., 2020; Xu and Liu, 2023; Chmiel et al., 2024) as well as syntactic source speech characteristics shining-through in interpreters' target speech (Alonso-Bacigalupe, 2010; Sandrelli, 2018; Collard et al., 2019). Shining-through observed at the syntactic level (Alonso-Bacigalupe, 2010; Sandrelli, 2018; Collard et al., 2019) is observed when the interpreting product is largely based on the external shape of the source speech and such "word-for-word translation in SI with structurally and lexically similar language pairs is not an unwanted accident nor is it due to low interpreters performance or quality, but rather the natural consequence of a global strategy of effort maximisation, the aim of which is to obtain maximum communicative efficiency with the minimum cognitive effort" (Alonso-Bacigalupe, 2010, p.51).

The idea of a "linearity constraint" of the interpreting process introduced by (Shlesinger, 1995) is clearly reflected in such patterns where it is cognitively easier not to change the structure where possible. Interpreter's adherence to linear constraints may not only have an effect on interpreter's syntactic patterns in the target, but also result in certain lexical patterns such as a higher proportion of function words (Liu and Dou, 2023). For more distant language combinations, interpreters may resort to other strategies, such as segmentation, in order to cope with the linearity constraint. This may lead to lower levels of syntactic complexity (Liu et al., 2023).

In order to prevent interference or shining-though of the source language, strategies such as chunking, restructuring, or reordering can be applied (Chmiel et al., 2024), however, potentially leading to an increase in cognitive load when interpreting asymmetric structures (Seeber, 2011). Chunking seems to be less effortful (Donato, 2003; Gile, 2009; Li, 2015) than other strategies such as reordering, which induces higher

cognitive load (Ma et al., 2020; Chmiel and Lijewska, 2019). Using shorter, simpler syntax may help interpreters speak more quickly and reduces the risk of errors under time pressure (Liu et al., 2023). While preparing terminology and background knowledge may help with lexical challenges, interpreters need to rely on their experience with source language structures to process syntactic structures of the source speech (Chmiel et al., 2024).

The findings for written language suggest that the results on lexical and syntactic complexity may differ with lexical and syntactic simplification not always going hand in hand (Liu et al., 2022). The overview of studies given in Section 2.4 shows that this is also the case for the spoken mode of translation.

In general, it can be said that syntactic simplification in interpreting is observed via shallow syntactic measures such as shorter sentences or embeddedness (Russo et al., 2012; Bernardini et al., 2016; Gast and Borges, 2023), more sophisticated computational measures on syntactic complexity (Liu et al., 2023; Bizzoni et al., 2020), and also via shorter dependency distances (Bizzoni et al., 2020; Xu and Liu, 2023; Chmiel et al., 2024). This thesis also uses the dependency locality framework (Gibson, 1998, 2000) for syntactic investigations, as it not only measures syntactic complexity, but also provides a link to cognition reflecting the memory and integration costs of each element in a sentence. Dependency locality measures can be used as a proxy of the processing difficulty of a sentence (Gibson, 2000; Liu, 2008). As with surprisal measures used to investigate lexical simplification, dependency locality measures correlate with reading times (Grodner and Gibson, 2005) and long distances between syntactically related words can lead to comprehension difficulties (Grodner and Gibson, 2005; Bartek et al., 2011). With a tendency of efficient encoding via dependency length minimisation found in many natural languages (Futrell et al., 2015), interpreters trying to use their cognitive resources in an optimal way may therefore also be expected to minimise dependency distances, if possible. However, as previous research shows, such pressures to optimise encoding of the target speech may be counterbalanced by source speech and language properties. This thesis aims to investigate an overall syntactic simplification effect for the language pair German and English, a language pair that has been understudied in the context of simultaneous interpreting. In addition to a general approach to simplification via a comparable dataset, potential triggers of simplification are investigated via a parallel approach, comparing target speech and the source input on which it was based.

2.5 Summary of theoretical background, hypothesis and definition of simplification

This section concludes the theoretic background by summarising the key findings of the literature review and defining the research questions (RQs). Furthermore, it lays out the definitions of simplification used in this thesis.

2.5.1 Summary of theoretical background, hypothesis and research questions

Simplification is found to be one of the specific properties that are known to characterise written translations (cf. Section 2.1.2). While some other translationese trends may reinforce simplification in translations, such as explicitly adding connectives (i.e. *explicitation*) (Blum-Kulka, 1986) and making the translation product clearer for the translation audience, other translationese tendencies such as *shining trough* (Teich, 2003) or *implicitation* (Klaudy and Károly, 2005) may lead to unexpected sequences or to a greater comprehension effort for the target audience who have to infer elements that were left implicit (cf. Section 2.1.2). In simultaneous interpreting, there is a particular pressure to manage cognitive resources efficiently in order to deal with imminent time constraints and not to exceed total cognitive capacity (Gile, 1997; Seeber, 2011) (cf. Section 2.2.1). Certain strategies are available to interpreters to deal with the cognitively challenging task of simultaneous interpreting with different implications for interpreter's cognitive load and *simplification* tendencies. Interpreting structures with *syntactic symmetry* (i.e. keeping the syntactic structure of the source if possible) is considered cognitively easier than other strategies (Seeber, 2011; Seeber and Kerzel, 2012b). The interpreting strategy of *transcoding* (Kalina, 1998) may lead to patterns in the target language that may be grammatically correct, but are potentially more unexpected than other translation solutions, with source language structure *shining-through*. *Syntactic transformation* (Kalina, 1998), on the other hand, can lead to less unexpected output, but requires more memory effort to intermediately store syntactic components in memory until they are used later on in the sentence. Reordering of source speech input for the target production is generally seen as more effortful (Defrancq and Plevoets, 2018). When interpreting syntactic asymmetric structures, interpreters may *wait* or *stall* (Seeber, 2011; Seeber and Kerzel, 2012b), increasing memory load in the former case and adding neutral fillers or repetitions in the latter case. The same strategy may lead to contrary implications for a potential simplification effect in the target product. Neutral fillers added would potentially be low surprisal items, while repetitions might lead to unexpected, high surprisal sequences. The strategy of *chunking* may be cognitively beneficial compared to *waiting* or *stalling* (Ma et al., 2020), resulting in shorter segments, possibly also shorter sentences overall, and therefore a more simplified output. On the semantic level, strategies such as *generalisation, summarisation, omission, approximation* or *semantic compression* may either simplify the target product if less semantic content is transferred, or in case of omitted content that is necessary for cohesive links of the target product, may make processing more difficult for the target audience. Generally, it can be said that, if the application of these interpreting strategies is automated, they relieve cognitive load (Li, 2015).

Research Questions: Interpretese – specific properties of interpreted language

As discussed in Section 2.1, translations differ from original written production in the same language. Given the complexity and particularly the cognitive implications of the task of simultaneous interpreting as well as the strategies applied by interpreters, we expect simultaneous interpreting to differ significantly from original production in the same language. As research has shown that interpretese is not only influenced by task inherent constraints, but other factors such as source speech properties (e.g. delivery rate or mode or lexical density) (Plevoets and Defrancq, 2020, 2016, 2018a,b), the language pair (Dayter, 2018) and strategies applied by the interpreter (Alonso-Bacigalupe, 2010; Ma et al., 2020) also play a role. The first RQs addressed in this thesis are, therefore, stated as follows:
RQ 1a: Does interpreting differ significantly from original speech in the same language?
RQ 1b: What are the linguistic features that contribute to the specific linguistic properties of language that interpreters use?

To address these research questions, this thesis focusses on an exploratory approach, using relative entropy as a data-driven method to detect distinctive differences between interpreting when compared to original spoken language and vice versa (Chapter 4). With the hypothesis that interpreting will differ significantly from original spoken language, language models for comparable subcorpora in English and German from EPIC-UdS are compared as to how similar or dissimilar original and interpreted language is. The method of relative entropy gives the advantage that no prior selection of features is required as the distinctive features, in our case words, are the result of the analysis and only significant differences are included in the results. If RQ 1a and the hypothesis that interpreting differs significantly from original speech in the same language is confirmed, it is assumed that *interpretese* is not characterised by the same linguistic features as *translationese* (cf. Section 2.1), due to major differences in the production process (cf. Section 2.2). RQ1b explores the specific properties that characterise *interpretese*. This thesis aims at giving a clearer description of the interpreting product and uses relative entropy to detect specific characteristics of interpretese.

Research Questions: Simplification in simultaneous interpreting

After first investigating general interpretese trends in RQ 1a and 1b, the following RQs use the results of the exploratory approach (Chapter 4) and zoom in on simplification as one particular aspect of interpretese. RQs 2 investigate lexical *simplification*, while RQs 3 focus on syntactic aspects of *simplification* (more specific RQs are stated in individual sections below). The general assumption on which this thesis is based is that due to the overall complexity of the task and the resulting cognitive implications, interpreters are required particularly to use the linguistic options available to them

to optimally encode the message. The tendency to encode messages efficiently can also be found in natural languages (Gibson et al., 2019), but it can be expected to be more pronounced in interpreting. This thesis investigates whether simplification trends can be observed in the interpreting product. Such simplification trends would be assumed to be the result of the pressure to optimally encode the message due to the general cognitively challenging production conditions in interpreting.

Corpus-based approaches to cognitive load in interpreting show that a sentence is a suitable unit of investigation as cognitive load does not seem to be transferred across sentences boundaries (Plevoets and Defrancq, 2020). Traces of cognitive load can be investigated through the use of corpora (Plevoets and Defrancq, 2020; Chmiel et al., 2023; Jiang and Jiang, 2020). It can be observed via corpus-based studies that high cognitive load in simultaneous interpreting is linked to source speech properties such as lexical density, formulaicity, and syntactic complexity, as well as target speech properties such as formulaicity and lexical density (Plevoets and Defrancq, 2020, 2016, 2018a,b). When interpreters perceive more input load, they seem to reduce output load, e.g. by producing more standard material (Plevoets and Defrancq, 2020). Therefore, it can be expected that means of simplification will be observed in generally effortful conditions.

Production processes make a significant or even dominant contribution to cognitive load in simultaneous interpreting (Defrancq, Bart, 2024). Production can be influenced by producing the most easily retrieved translation solution. This may include automatic or ready-made translations solutions (Alonso-Bacigalupe, 2010; Setton, 1999), including formulaic sequences (Kajzer-Wietrzny and Grabowski, 2021), linearity (Shlesinger, 1995; Alonso-Bacigalupe, 2010) or chunking (Ma et al., 2020). Generally, simplification is a multifaceted phenomenon (Xiao and Dai, 2014; Liu et al., 2022) and it has to be investigated as to what extent factors of the source language and speech may counterbalance a potential tendency to optimally encode the target message. After all, source speech properties cannot be influenced by the interpreter. Within the interpreting process, the only way to influence their own cognitive load is by adapting target speech production and potentially simplifying at various linguistic levels. While Plevoets and Defrancq (2020) observe interpreters to "balance comprehension load by reducing production load" (Plevoets and Defrancq, 2020, p.30) and observe such simplification in local lexical density and formulaicity, this thesis investigates a general simplification trend.

Traditional corpus-based translationese measures on simplification give diverging results for simultaneous interpreting (cf. Section 2.4.2), while measures used in readability research show simplification in simultaneous interpreting (Kunilovskaya et al., 2023a). In addition to traditional measures, this thesis employs corpus-based measures that are known to provide a direct link to cognition, i.e. *surprisal* (Hale, 2001) for lexical investigations, and dependency locality measures (Gibson, 2000) to approach syntactic simplification.

Research Questions 2: Lexical simplification Previous research on lexical simplification in simultaneous interpreting gives a mixed picture (cf. Section 2.4.2). Information-theoretic measures and surprisal in particular have been shown to be useful in translation process research (Carl, 2021; Wei, 2022; Lim et al., 2023) and product analysis (Rubino et al., 2016; Lapshinova-Koltunski et al., 2022; Kunilovskaya et al., 2023b). Surprisal not only considers lexical frequency, but also considers lexical use in context and provides a clear link to cognition (Hale, 2001; Teich et al., 2020). As mentioned above, cognitively challenging conditions in simultaneous interpreting are expected to lead to expected sequences in the interpreting product, which may potentially be counterbalanced by properties of the source speech shining-through, such as the use of cognates or unexpected translation solutions for unexpected source speech items.

Processing times for linguistic input can generally be linked to lexical frequency (Brysbaert et al., 2018; Weber and Crocker, 2012). Lexical frequency is also particularly linked to cognitive load in interpreting (Plevoets and Defrancq, 2020; Chmiel et al., 2023). A particular processing advantage can be found for high frequency cognates, which can be retrieved particularly quickly and interpreters may use to reduce cognitive load (Chmiel et al., 2023). However, the use of cognates is source speech triggered and may lead to unexpected (high surprisal) words in context. Investigating lexical simplification via surprisal covers frequency and context aspects while at the same time linking it to cognitive processing in interpreting.

When investigating simplification at word level, a particular distinction has to be made between function words and content words as the main carriers of the message to be transferred. While function words are generally accessed faster than content words, access of content words, again, is known to depend on word frequency (Segalowitz and Lane, 2000). Within the different types of content words, nouns seem to slow down production more than verbs (Seifart et al., 2018). A more detailed investigation of different lexical POS is to shed light on particular simplification tendencies of these linguistic categories. To investigate the hypothesis of lexical simplification in simultaneous interpreting, the following research questions are addressed:
RQ 2a: Is simultaneous interpreting more simplified at the lexical level than comparable originals in the same language?
RQ 2b: What are the particular triggers of lexical simplification in the source speech input? This question focusses on distinctive POS for interpreting which are mainly content words as the main carrier of a message.

RQ 2a uses a comparable corpus approach, comparing simultaneous interpreting to original speech in the same language (English and German). While traditional corpus-based measures on simplification give mixed results for other language combinations (Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2015; Dayter, 2018; Ferraresi et al., 2018), selected measures based on the results of the exploratory analysis (Chapter 4) are applied to the English and German comparable subcorpora, acknowledging that German interpreting is not much studied in the literature. The approach taken is

uses only traditional measures that are linked to distinctive features of interpreting when compared to original speech. This is investigated in Section 5.1. Section 5.2 investigates RQ 2a and RQ 2b via surprisal, linking frequency in context to cognitive processing. The comparable corpus approach (RQ 2a) for German and English comparable data (German interpreting from English sources and English interpreting from German sources) is supplemented with an additional English dataset for English interpreting based on Spanish source speeches. This allows for verifying whether trends found in interpreting are linked to the language combination German-English or can also be found as source language independent trend. RQ 2b takes a parallel perspective and looks at high and low surprisal words in interpreting and investigates their source speech triggers.

Particularly relevant for lexical aspects, as discussed in the background discussed in this chapter, are certain interpreting strategies. When stalling, interpreter's may add neutral fillers that are assumed to be expected (and low surprisal) words. In case of added repetitions, this could also lead to unexpected word sequences in context. Furthermore, it can be expected that interpreter's lexical production is influenced by source text lexis, and a cognate facilitation effect (Chmiel et al., 2021) may lead to unexpected lexical use in the target speech (i.e. high surprisal items). These research questions on lexical simplification are addressed in Chapter 5.

Research Questions 3: Syntactic simplification While surprisal can be used to investigate lexical choice linked to the cognitively demanding conditions in simultaneous interpreting, a complementary approach to sentence processing and resulting effects on the syntactic level is Dependency Locality Theory (DLT) and Dependency Locality Minimisation (DLM), particularly explaining storage and integration costs in spoken language production (Collins, 2014; Konieczny, 2000). DLM is generally especially related to syntactic choice, where the language producer has different options to communicate the same message (Hawkins, 1995, 2004; Temperley, 2007, 2008; Liu et al., 2017; Ferreira and Dell, 2000; Arnold et al., 2000).

In spoken language, speakers are expected to pursue "their own communication goals or according to availability-based production accounts, they might be realising the most readily available constituents" (Rajkumar et al., 2016, p.225), rather than actively seeking to facilitate comprehension and adjusting their speech to the needs of the listeners. In simultaneous interpreting, the use of shorter dependency length is directly related to cognitive processing (Liang et al., 2017) and shared dependency structures in source and target speech are found to be cognitively beneficial (Jiang and Jiang, 2020).

As there is no empirical evidence that cognitive load in simultaneous interpreting is imported to following sentences (Plevoets and Defrancq, 2020; Chmiel et al., 2023) and interpreters seem to be able "to reset their working memories at the sentence boundary" of their own production (Plevoets and Defrancq, 2020, p.35), it is

important to note that at the end of the sentence produced by the interpreter, all dependency relations in interpreter's target speech are resolved and no further memory effort is required to store open dependency relations related to that sentence.

Although the general motivation in simultaneous interpreting to encode the message efficiently and use a simpler syntactic structure, because of the reasons laid out above, a trend to use shorter dependencies could be potentially counterbalanced by a source language influence. Generally in multilingual communication, longer dependency distances are observed (Fan and Jiang, 2019; Liang and Sang, 2022; Wang and Liu, 2013) and syntactic priming may also play a role (Pickering and Branigan, 1999). Furthermore, sentence-by-sentence processing in simultaneous interpreting may impact syntactic structure (Liang et al., 2017). Nevertheless, several studies report shorter dependency distances in simultaneous interpreting, although some of the results may be due to interpreters, in these studies, working into their non-native language (Liu et al., 2023; Xu and Liu, 2023).

As this thesis uses English and German interpreting and original speeches, it is particularly relevant to consider that general dependency length minimisation trends differ between these two languages, with stronger minimisation trends in English with a mainly right-branching sentence structure (Temperley, 2007) compared to German as head-final language (Gildea and Temperley, 2007; Park and Levy, 2009; Futrell et al., 2015; Dyer, 2023). Furthermore, DLM tends to be more pronounced in long sentences (Futrell et al., 2015; Rajkumar et al., 2016). Although sentence length is generally shorter in simultaneous interpreting than original spoken data (Russo et al., 2012; Bernardini et al., 2016), long sentences in simultaneous interpreting can be expected to be influenced by the source language input. Long dependency length in the source are known to increase cognitive load for the interpreters (Jiang and Jiang, 2020).

Before addressing syntactic complexity via a dependency length analysis, this thesis first takes a general approach on simplification and looks into sentence length in the comparable and parallel corpus data as well as investigates how interpreters deal with source input at the sentence level. Specifically, the following research questions are asked:

RQ 3a: Can syntactic simplification measured via sentence length (comparable and parallel approach) and alignment analysis (parallel approach) be observed in interpreting?

Furthermore, considering the aspects for and against potential dependency length minimisation, the following research questions are asked to give a more detailed insight about syntactic complexity.

RQ 3b: Can syntactic simplification measured via dependency length minimisation be observed in the interpreting product (comparable approach, comparing originals and interpreting in the same language)?

RQ 3c: What are the triggers of a potential minimisation effect? Does shining-through of source language structure counterbalance a possible minimisation effect

(parallel approach, comparing source and target speech)?

This thesis complements previous (experimental) studies showing the relation between cognitive load and interpreting by investigating a general trend of simplification in simultaneous interpreting and in particular adding empirical knowledge about the language pair German-English.

2.5.2 Definition and operationalisation of simplification

The basic definition of simplification used in this thesis is taken from the translation literature, where simplification is defined as the process of translators to "subconsciously simplify the language or message or both" (Baker, 1996, p.176) in their target product. This thesis focusses on investigating the interpreting product with the expectation to find simpler target speech compared to original spoken language. It therefore does not investigate the interpreting process and related cognitive load directly, but rather investigates implications of the particular language production conditions in simultaneous interpreting that would be visible in the interpreting product. It investigates the result of the simplification process, which is "making things easier for the reader" (Baker, 1996, p.182) or, in case of spoken translation, easier for the listener. It is important to note that this only refers to the language use of interpreters at the lexical and syntactic levels and does not include prosodic aspects of the interpreting product that certainly also have an effect on recipients perception.

Baker's definition of simplification for the (written) translation process can generally also be applied to spoken translation. Regarding simplification as a "subconscious" process, it has been established in Section 2.2.2 that the strategies that can be applied by interpreters generally refer to automated processes rather than conscious decisions to apply a certain tool. Simpler use of the language and/or simplifying the message of the source during the translation process can be related as follows for the process of simultaneous interpreting: Simplifying the message of the source speech relates to the semantic level where interpreting strategies such as *generalisation, summarisation* or *omission* may lead to simpler target speech.

This thesis focusses on language use particularly related to lexical and syntactic aspects. Lexical simplification has traditionally been defined as "the process and/or result of making do with less words" (Blum and Levenston, 1978, p.399). "Less words" can refer to reduced lexical variation in linguistic use with interpreters using fewer different words, making the interpreting product less varied and easier for the target audience to process. Furthermore, it also refers to the syntactic level, and particularly to shorter sentences, a less complex syntactic structure, or shorter dependency distances.

Regarding the lexical level, simplification has traditionally been defined as texts or speeches with lower lexical density (i.e. lower information load), lower lexical variety and resolved ambiguities, (Baker, 1996; Laviosa, 1998; Dayter, 2018). Lower lexi-

cal variety generally refers to lower type-token ratio, narrower range of vocabulary used, higher proportion of high-frequency words, increased level of repetitiveness, and lower degree of lexical sophistication (Laviosa, 1998; Kajzer-Wietrzny, 2015), as well as lower occurrences of hapax-legomena (Lv and Liang, 2019). It includes the use of superordinate terms and synonyms, basic-level category terms, concept approximation, the use circumlocutions and paraphrase, and transfer of all the functions of the source language word to the target language equivalent (Laviosa, 1998).

At the syntactic level, simplification means breaking up long sentences into shorter ones during the translation process (Baker, 1996), which is sometimes also referred to as *stylistic simplification* (Laviosa, 1998). Reduced syntactic complexity can be observed via the use of finite instead of non-finite structures (Baker, 1996; Laviosa, 1998), but also by the use of short instead of elaborate collocations, reducing repetitions and redundancies, shortening circumlocutions and leaving out modifiers (Laviosa, 1998).

The simplification definition used in this thesis takes an even broader angle, including other trends observed in translations that would make the interpreting product easier for the recipient to process, and thus making the interpreting product more simple. This includes explicitly spelling things out rather than leaving them implicit (Baker, 1996), such as explicitly using cohesive links or replacing a general lexical word of the source with a more specific word in the target, reducing ambiguity (Toury, 1995). It also incudes Kuo's definition of simplification that sees texts with increased redundancy and amplified explanations as more simplified (Kuo, 1993 in Crossley et al., 2007), although the latter aspect may not play a role in simultaneous interpreting with the imminent time pressure.

While these aspects of such a broad definition of simplification are not particularly investigated in this thesis, it has to be mentioned that further translationese trends may also lead to more simplification. Normalisation may lead to high frequencies of certain lexical patterns, the use of common lexis instead of unusual lexis, or preference of standard language over dialect. The results of such processes may make the interpreting product more expected in context, i.e. leading to lower surprisal.

This thesis therefore employs a wider definition of simplification, capturing the general notion of "making things easier for the reader" (Baker, 1996, p.182) and potentially also for the producer, which may include explicitation and normalisation tendencies.

Lexical simplification in this thesis is thus defined and studied as follows. More expected words in context are considered more simplified, i.e. a high degree of lexical simplification is indicated by the use of low surprisal words on average, whereas unexpected words in context can be linked to a low degree of simplification. This is operationalised via the information-theoretical measure of surprisal, showing how predictable a linguistic item is in a given (variable) context. As for syntactic simplification, simplification is operationalised by the use of measures of dependency grammars, particularly the distance of individual syntactically related words in a sentence to approximate the required effort for maintaining a linguistic item in working mem-

ory. Using these more process-related measures is hoped to give a more comprehensive picture of process-related traces that can be observed in the translations product.

Part II

Data and Methodology

Chapter 3

Corpus and methodological design

As described in Section 2.4, what we empirically know about interpretese is based on relatively few interpreting corpora, covering a limited number of languages. With few exceptions, the available interpreting corpora are situated in the political field, e.g. SIREN for English-Russian (Dayter, 2018), CEPIC for Chinese-English (Pan, 2019), and many of them in the context of the European Parliament (EP) such as EPIC covering English, Italian and Spanish (Russo et al., 2012), EPICG covering French, English, Dutch and German are source and target languages and Spanish as Spanish as a source language (Defrancq, 2015), TIC with English as target language based on French, Spanish, German and Dutch source speeches (Kajzer-Wietrzny, 2012) PINC for Polish-English (Chmiel et al., 2021), and ESIC for English sources interpreted into Czech and German (Macháček et al., 2021). Interpreting corpora covering other genres are rare, e.g. HeiCIC covering German, English, French, Italian, Spanish, Portuguese, Russian and Japanese (Kunz et al., 2021), NAIST for English-Japanese (Doi et al., 2021), or BST for Chinese-English (Zhang et al., 2021)). Refer to Przybyl et al. (2022b) for a more detailed overview. All of these simultaneous interpreting corpora cover English as a source or target language. However, German is underrepresented in the available corpus resources, covering only a small proportion of the available corpus data.

The political setting and in particular the EP provide many advantages for interpreting corpus compilation. Besides the most obvious advantage of freely available interpreting and original spoken data in all official European languages, available as video and audio recordings, there are other factors that make EP data very valuable. The setting is highly standardised across the individual speeches, including strict application of the allocated speaking time (Seeber, 2017). The interpreters work in highly professional working conditions with professional equipment, a good view of the speaker, taking regular turns with their booth colleagues, and experience within this field of work, just to name a few examples.

On the downside, it has to be mentioned that the EP is a very particular type of

setting for conference interpreting with various factors different from other conference settings. Therefore, the possibility of generalisation of findings with this particular setting is limited. Such differences include the professional setting including equipment, preparation, and view of the speaker which varies greatly in the interpreting market. Individual speeches in the EP are generally shorter and faster than in other settings (Russo et al., 2012). However, one of the biggest differences is probably the level of preparedness regarding the knowledge of situation, context, and content of the EP interpreters compared to the freelance conference interpreting market. EP interpreters are mainly permanent interpreters as well as accredited freelance interpreters who usually work for the EU institutions on a regular basis. Compared to the freelance interpreting market outside the EU institutions, freelance conference interpreters generally work for many different clients and have to adapt and prepare for diverse conference and business settings. Such differences have to be considered when drawing conclusions from EP corpus data and relating them to simultaneous interpreting in general. Nevertheless, advantages of using European Parliament interpreting data outweigh the disadvantages by large, and alternatives regarding availability of simultaneous interpreting data in sufficient amounts are rare.

Adding a EP interpreting corpus resource with German-English as primary language pair aims to fill the void of German being under-represented in this context so far. Furthermore, the comparability of the EPIC-UdS data with other corpora of the EPIC family provides great potential for studies beyond the scope of this thesis. In the present thesis, the German and English subsets of the EPIC UdS corpus as described in Przybyl et al. (2022b) and additional corpus variants are used as described in the following section.

3.1 EPIC UdS

The studies carried out in the context of the thesis are based on the European Parliament Interpreting Corpus at Saarland University (EPIC-UdS). EPIC UdS belongs to the European Parliament Interpreting Corpora family and is comparable in terms of corpus design and metadata collected to other EPIC corpora (EPIC (Russo et al., 2012), EPICG (Defrancq, 2015)). The English dataset (English original speeches and English simultaneous interpreting with German and Spanish as source languages) is mainly based on transcripts drawn from TIC (Kajzer-Wietrzny, 2012) and some English original speeches are taken from EPICG (Defrancq, 2015). The German transcripts were produced at Saarland University. All newly transcribed as well as the revised existing transcripts follow the EPICG transcription guidelines (Bernardini et al., 2018) with minor modifications [1] to ensure comparability across the datasets. These guidelines specify that spoken language characteristics, such as filled pauses, false

[1] Annotation of sentences and insertion of punctuation are not specified in these transcription guidelines and are described in Section 3.1.2.

starts, and truncated words, are orthographically transcribed (cf. Table 3.1). The transcripts are based on videos of the European Parliament Proceedings published by the European Parliament[2]. Speeches were held from 2008 to 2013 by members of the European Parliament (MEPs).

Feature	Transcription
Silent pause	/
Filled pause	euh, hm, hum
Mid-word pause	spea/ euh ker [speaker]
Rising intonation	[?]
Non-verbalized noise	[noise], [breath]
Non-standard pronunciation	report [repo:rt]
Inaudible segment	[inaudible]
Mispronunciation	plemary [plenary]
Truncated word	propo/
Ambiguity	they [?there]
Sentence-equivalent unit	↲

Table 3.1: EPIC UdS transcription for interactional and non-verbal acoustic features based on EPICG.

Rich metadata are available for source and target speeches (i.e. originals and their interpretation) in accordance with metadata collected for EPIC and EPICG (Bernardini et al., 2018) (cf. Table 3.2). Speaker-related metadata, as well as some speech event-related metadata such as *general topic* and *date*, are taken directly from the EP website. Speaker's *nationality* is used to operationalise the native language as English is spoken frequently by non-English native speakers in the plenary sessions. In the corpus, only British and Irish nationals speaking English in the EP are included for English original speeches. The English original subcorpus covers speeches by 56 different speakers. For German original speeches, only German and Austrian nationals giving their speech in German are included, covering 87 different speakers in the German original subcorpus[3]. This, of course, can only be considered as an approximation of the native language status.

Metadata for interpreters cannot be collected easily; however, some assumptions can be made: *gender* information can be extracted from the audio file, but with limited reliability. By selecting interventions from different dates/years, it can be assumed that the interpreted speeches were rendered by different interpreters. Additionally, interpreters' voices were manually identified by listening to the audio file,

[2]downloaded from https://www.europarl.europa.eu/plenary/en/debates-video.html (accessed 2018/2019)

[3]All speaker related metadata were anonymised in the corpus and only collected to make sure that the corpus covers a variety of speakers and study results are not only idiosyncratic preferences.

giving each voice an interpreter id. This identification is also not completely reliable, but combining both approaches, it can be assumed that between 20 and 30 different interpreters were delivering the interpreted speeches in each subcorpus (cf. Table 3.3). Furthermore, it can be assumed that all interpreters working in the European Parliament are trained professionals, as all interpreters in the European Parliament have to pass an accreditation test, which can only be taken by graduates of Master programmes for Conference Interpreting or equivalent programmes. At least for the bigger European languages, interpreters generally work into their A-language (mother tongue). This particularly applies to the English and German booths. Therefore, we assume that speakers as well as interpreters within EPIC UdS are native speakers. Furthermore, it was manually verified that ear-voice span (i.e. the time difference between the original speaker's utterance and the interpreter's rendition of this utterance in the target, also referred to as lag, or décalage) is not consistently larger than 4 seconds. Thus, the interpretations in the corpus are defined as direct interpretations, involving no third "pivot" language (Bernardini et al., 2018).

Speaker related	Name Gender Nationality (operationalised for native language)
Interpreter related	Gender Native language
Speech event related	General topic (as indicated by EP) Title of debate (as indicated by EP) Date Length in words Length in seconds Delivery rate: in words per minute (w/m) Delivery rate: slow, medium, high Source text delivery type: read, impromptu, mixed

Table 3.2: EPIC UdS metadata.

More speech-event-related metadata were collected using the transcripts and audio files of the individual speeches. Transcripts and audio files are used as the basis for *length* of the speech *in words* and *in seconds*, or the combination of the two as *delivery rate in words per minute*. The assigned classification for *delivery rate as slow, medium or high* is chosen specifically to reflect the context of European Parliament plenary sessions, where MEPs have to respect very strict rules for the allocation of speaking time and therefore often prepare their speeches, generally speak very fast and try to fit as much in as possible (Monti et al., 2005). The classification of the delivery rate as *slow* $\leq$130w/m; *medium* = 131-160w/m; *high* $\geq$161w/m has been set in accordance with EPIC (Bernardini et al., 2018). In other interpreting settings,

different classifications would apply, e.g. in the Italian conference interpreting market, a speed of delivery of 150w/m would be fast (Monti et al., 2005). Pöchhacker considers 100-120 w/m as "comfortable" speech of delivery for interpreting (Pöchhacker, 2004, 129).

The level of preparedness is captured as *source text delivery type*. The video recordings were used to see whether the original speaker was holding a script and reading from it (*read*) or whether they were speaking without a script (*impromptu*). If the speaker was holding a script, but only reading from it occasionally, the speech was classified as *mixed* delivery type. No information on the preparedness or availability of speakers manuscripts for interpreters was available. However, due to the short length of the interventions and the large number of speakers in a plenary session, we expect interpreters not to be able to produce an interpretation with text. Interpreted speeches are expected to be freely delivered and not based on prepared translations.

	Tokens	Sentences	Speeches	Running time	Individual speakers
EPIC UdS EN					
ORG EN	67,526	3,622	137	6,65 hours	52
SI DE EN	58,503	3,623	165	6,61 hours	at least 28
SI ES EN	54,630	3,076	163	5,64 hours	at least 20
EPIC UdS DE					
ORG DE	56,488	3,409	165	6,61 hours	87
SI EN DE	57,532	4,076	137	6,68 hours	at least 27

Table 3.3: Corpus overview EPIC UdS V2, original spoken English (ORG EN) (ORG EN), original spoken German (ORG DE) (ORG DE), simultaneous interpreting from German into English (SI DE EN) (SI DE EN), simultaneous interpreting from Spanish into English (SI ES EN) (SI ES EN), simultaneous interpreting from English into German (SI EN DE) (SI EN DE).

Table 3.3 gives an overview of the corpus size for EPIC UdS V2 in the number of tokens, sentences (cf. Section 3.1.2) and speeches for each subcorpus, as well as the total length of the recordings. The following abriviation for the individual subcorpora are used in this thesis: original spoken English (ORG EN), original spoken German (ORG DE), simultaneous interpreting from German into English (SI DE EN), simultaneous interpreting from Spanish into English (SI ES EN), and simultaneous interpreting from English into German (SI EN DE). The number of tokens includes filled pauses, truncated words, and other transcribed spoken features, but does not contain punctuation marks as they are not included in the transcription. This corpus variant is the basis for the exploratory investigations in Chapter 4, as well as the lexical investigations using traditional corpus-based measures (Section 5.1). The number of

tokens and the level of annotation differ slightly in the different corpus variants as the amount of spoken language features differs (POS correction for filled pauses in EPIC UdS V3 for better parsing accuracy and exclusion of most spoken features such as filled pauses and truncated words in corpus variant EPIC UdS (aligned, parsed, surprisal)). EPIC UdS (aligned, parsed, surprisal) is used to investigate expectedness in context (Section 5.2.2) and EPIC UdS V3 for the syntactic investigations (Chapter 6).

Corpus version	Basis for study
EPIC UdS V2	KLD analysis in Chapter 4
EPIC UdS V3	Traditional measures in Chapter 5 (5.1)
EPIC UdS (aligned, parsed, surprisal)	Surprisal analysis in Chapter 5 (5.2.2)
EPIC UdS V3	DL analysis in Chapter 6

Table 3.4: EPIC UdS corpus variants and their use in this thesis.

3.1.1 Surprisal

The lexical investigations in Section 5.2.2 use the corpus version EPIC UdS (aligned, parsed, surprisal). Each word in this corpus variant is annotated with a surprisal value (cf. Section 5.2.1 for more information on the methodology used). The language models used as basis for calculating surprisal for this corpus variant are based on the combined EPIC UdS original and interpreting corpus for each language, i.e. English original speeches and English interpreting from German and Spanish as source for the English dataset, and German originals and German interpreted speeches from English sources for the German dataset. This approach is used to approximate the exposure of European Parliament speakers and interpreters to the language spoken and heard in this setting (i.e. original speeches as well as interpreted language). Other training options to calculate surprisal such as using more data (e.g. the written Europarl corpus (Koehn, 2005) or surprisal retrieved from large language models) would not reflect the particularities of the spoken and interpreted characteristic of the dataset at hand.

$$S(w_i) = -log_2 p(w_i|(w_{i-1}w_{i-2}w_{i-3})) \tag{3.1}$$

When calculating surprisal for a speech, the respective speech was excluded from the training set. Surprisal is calculated on a 4-gram basis (cf. Equation 5.3), using the context of 3 previous tokens to estimate (un)predictability of a given word. The approach respects (artificially inserted) sentence boundaries based on sentence-equivalent segments (cf. Section 3.1.2). Hapax legomena were replaced with a placeholder string. Words are defined as tokenised strings, which means that different word forms of the same lemma are treated as individual words. More detailed information

is given in Section 5.2.1. A similar modelling approach was also used in Kunilovskaya et al. (2023b), however, instead of using all spoken data available for that language in the corpus as in the approach used in EPIC UdS (aligned, parsed, surprisal), the authors use a balanced dataset for each language. This approach considers the fact that especially for the English part of the corpus, the interpreting subcorpora combined (interpreting into English with German and Spanish as source) are much larger than the original spoken English subcorpus. If research question 1a on interpretese was confirmed and originals and interpreted speech in the same language differ significantly, an unbalanced dataset used for language modelling may affect the reliability of the surprisal estimates obtained via this approach. However, as the general trends observed in (Kunilovskaya et al., 2023b) are in line with the results found in this thesis, the approach taken here is considered reliable.

3.1.2 Sentence annotation

As the transcription standards for EPICG that were used as the basis for transcribing EPIC UdS do not specify guidelines for dealing with sentence boundaries, sentence segmentation has to be defined in more detail, particularly since the segmentation approach used directly influences the results of the syntactic investigation (Chapter 6). The following sections describe in more detail the approach used for EPIC UdS.

Automatic annotation of text corpora requires segmentation, e.g. in order for parsers to detect sentence boundaries and yield better results. Furthermore, for aligning source and target texts beyond the text level, a more fine-grained segmentation is necessary. In written texts, the required segmentation is given by punctuation marks or digitally via line breaks, indicating sentences. Spoken language corpora usually do not include punctuation marks. EPICG uses pauses as boundaries for segmentation (Bernardini et al., 2018). In prepared and read speeches, pauses and pause length might be a valid marker to define clauses or units (Henderson et al., 1965). In interpreted speech, however, the end of an independent clause is not always characterised by a pause. In contrast, pauses can occur triggered by the source language input (interpreters need to wait for more input in order to understand the meaning) or due to processing activities (especially clear when pauses occur in the middle of a compound noun, see examples 15 and 16)[4].

(15) ...große öffentliche *Ausgaben / programme* zum Beispiel im Wohnungsbau...

(16) ...denn es bildet die *Platt / euh form*...

EPIC uses pauses and other prosodic features to determine sentence-like units (Bernardini et al., 2018). However, pauses and pitch range differ in interpreted and

[4]As defined in the transcription guidelines, silent pauses are indicated by / , filled pauses by *euh*.

original speech (Christodoulides, 2013), with interpreted speech having a narrower pitch range than original speakers. The end of a clause in interpreted speech is often not characterised by a clear falling tone. Pauses do not always occur at the end of a clause. Using only prosodic features to determine the end of a clause can, therefore, be misleading. Without the audio file as reference, it is also not always clear how to set the sentence boundaries correctly. Example 17 shows a pause after a possible end of a sentence (*if Russia were to refuse*) and a further pause after a second possible sentence end (to withdraw troops). Segmentation in 17a. as well as 17b. are both possible. Only by listening to the audio file with a clear falling tone can it be decided that 17b. is the correct segmentation.

(17) ...the question then arises what we do / if Russia were to refuse / to withdraw troops / we are told we must have dialogue...

a. ...the question then arises what we do / if Russia were to refuse / ↲
to withdraw troops / we are told we must have dialogue...

b. ...the question then arises what we do / if Russia were to refuse / to withdraw troops / ↲
we are told we must have dialogue...

Although sentences might not be the best solution for segmenting spoken language (Pietrandrea et al., 2014), a sentence-based approach is determined to be suitable for monologic speeches (no turns of different speakers), which is the case for EPIC UdS data. However, a combined approach of listening to the audio file and using syntactic information to identify sentences is still not straight forward, as several main clauses are often joined by a coordinating conjunction, creating long parataxis of seven or more main clauses, without clear prosodic indication where to end the segment. As the aim is to align original speech and interpreters output as detailed as possible on segment level as well as get reliable parsing results, an approach to segment the smallest possible sentence unit was considered most suitable for this purpose (see example 18 with four individual segments). A line break is used to mark the end of a segment or sentence equivalent.

(18) it's obvious for a country ruled by the state of law ↲
and I think it's a great shame that we have to talk about this / as if it weren't obvious / ↲
and therefore / the fact that we're saying / there should be a ceasefire / stop this weaponry / ↲
and we should be able to do everything possible / in a humanitarian way to cope with this conflict / ↲

An exception was made to this approach to deal with paratactic elliptical sentences. A series of elliptical main clauses was not split into serval segments if the constituent that was left out was only included in the first main clause. In example 19, the second main clause misses *can be*, which is included in the first main clause. Therefore, the two clauses were not split. The third main clause is complete and was therefore segmented separately. For hypotaxis, the main and dependent clauses were kept in one segment (example 20). Furthermore, examples 21 and 22 are seen as sentence equivalents and segmented separately.

(19) so VAT free purchases can be sold on / the VAT pocketed ↲
and then the trader vanishes

(20) if / a dialogue / is a prerequisite for / peace / then a rejection of that dialogue / means a perpetuation of / warfare or of the conflict

(21) yes thank you President / Commissioner / colleagues /

(22) Dankeschön euh / Frau Präsidentin

In summary, the following approach was used for sentence segmentation in EPIC UdS to achieve good parsing results and facilitate source-to-target alignment:

- Comparing transcript and audio file
- Identifying main clauses in accordance with prosodic features (pause, falling tone) and syntactic information (mainly subject, verb)
- Segmenting main clauses separately / avoiding parataxis in one segment
- Segmenting the smallest possible sentence equivalent

3.1.3 Sentence alignment

In a next step, the transcripts of source and target speeches with sentence equivalent segments were aligned at the sentence level using the alignment component in Trados Studio 2021. Trados Studio proposes an automatic sentence alignment tool based on an internal confidence rating. The tool is designed for written translations that have distinct differences compared to the spoken transcripts at hand. Many of the rules for sentence alignment used by the software, such as matching numbers or acronyms in source and target, are orthographically transcribed in the spoken data. Therefore, the given confidence rating was ignored and all automatically aligned sentences were manually corrected to obtain reliable results. The tool allows for 1:1 alignment of

segments, where one source sentence is aligned to one target sentence, as well as 1:n and n:1 alignments with $n \geq 0$ and $n \leq 5$, where one source sentence is aligned to up to 5 target sentences or up to 5 source sentences being aligned to one target sentence. This limitation of the software was sufficient in most cases. However, the data also contain cases of 1:7 alignments, which were manually annotated.

	source	target
1:1	das ist ein perverses System ↲	and that's a perverse system ↲
1:1	das berührt den Dollar überhaupt nicht / weil nämlich die Wirtschaftspoli/ tik Kaliforniens im einheitlichen Währungsraum der / USA auf/ geht ↲	and / the economic policy of / California can't cope ↲
1:2	at the outset / let me put on the official record of the house / my thanks to all of my colleagues / both those who support / and those who oppose / the proposal that I'm putting forward / for their contributions for their input and in particular for their helpful advice and guidance along the road ↲	ich möchte / zuerst einmal / allen Kollegen danken auch für das Protokoll danken / die meine Vorschläge unterstützt haben ↲
		ich möchte ihnen danken für ihr Beiträge aber vor allem auch für ihre Ratschläge ↲
1:0	and we don't talk about it ↲	

Table 3.5: Alignment examples for source-to-target.

The alignment does not seek to reflect interpreting accuracy, but instead tries to link the target segments with the segments on which they were based. In some cases, this may be translations that are very similar to their source on the morphosyntactic level (cf. first 1:1 example in Table 3.5). However, this may also be translations that are rather a summary or an interpretation of the source, as in the second 1:1 example in Table 3.5, where the theme of both sentences is the same (*die Wirtschaftspolitik Kaliforniens - the economic policy of California*), however, the further content of the sentences differs. In some alignments, only part of the source input was translated as in the 1:2 example (Table 3.5). In this case, the long source sentence was split into two shorter target sentences. However, some parts of the source sentences (*at the outset*, noun modifications *helpful, along the road* and two of the four nouns at the end of the source segment) were left out. If no target segment was found that covered at least part of the source content, or vice versa, the sentence was defined as 1:0 alignment or 0:1 alignments, respectively (cf. 1:0 alignment in Table 3.5). Using

this approach, all segments were manually aligned.

3.1.4 Universal POS and dependency parsing

EPIC UdS V3 is a parsed corpus variant, using the universal dependency scheme to show dependency relations. This corpus variant is used as the basis for the syntactic analysis of simplification in Chapter 6. Opting for the smallest possible sentence unit as the segmentation approach as described in Section 3.1.2 is not only favourable for good source-to-target alignment, but also increases dependency parsing accuracy.

The universal dependency scheme annotates dependency relations between syntactically dependent words in a sentence and uses primarily relations between content words (de Marneffe et al., 2014). Function words are only used as direct dependents of the most closely related content words. Content words function as heads, which is supposed to maximise parallelism between languages, as content words vary less between languages than function words (de Marneffe et al., 2014). This content word-centric approach, besides being designed to work across languages, is favourable for the analysis of interpreting data as the interpreting process can be described as content transfer from one language (source speech) to another (interpreted speech).

In order to obtain the best possible accuracy for head-dependent relations and correct relation labels, some spoken language features, such as silent pauses and truncated words, were excluded from this corpus variant. The transcripts were processed segment by segment with Stanza tools (v1.2.3) and the corresponding language models (EN: EWT v1.2.3; DE: GSD v1.2.3; ES: AnCora v1.2.3). These language models were the default packages for each language at the time of parsing, pre-trained on the largest available treebank for that language (Qi et al., 2020). However, in order to improve accuracy, POS annotation for filled pauses (transcribed as *hum, hm* and *euh*) was manually corrected to POS tag *INTJ* to get a more consistent parsing result.

Corpus data for EPIC UdS V3 are encoded in CoNLL-U format and provides universal POS tags, fine-grained language-specific POS, and universal dependency relations. The CoNLL-U format is used for calculating dependency length measures in Chapter 6, whereas the universal dependency POS from the cqp web encoded version of EPIC UdS V3 serves as a basis for studying traditional measures of simplification. A list of dependency relations used in EPIC UdS V3 (parsing result of the Stanza tools used) is given in Appendix A.9.

Stanza model	LAS v1.0	UAS v1.0	LAS v1.3	UAS v1.3
EN EWT	83.59	86.22	84.91	87.41
DE GSD	80.61	85.39	80.66	85.29

Table 3.6: Parsing accuracy evaluation: stated for version 1.0 and 1.3.

Table 3.6 below shows the Stanza model's stated accuracy evaluation for Unlabeled Attachment Score (UAS, percentage of words with correct head) and Labeled

Attachment Score (LAS, percentage of words with the correct head and label) for the models v1.0. and v1.3. As there is no accuracy evaluation given for the models used (v1.2.3), we assume that the accuracy range of the models is between the values given in Table 3.6.

Table 3.7 shows the accuracy evaluation that was established for EPIC UdS V3. It is the result of a manual evaluation of the correct dependency label and the relation to the correct head for a subset of 50 randomly selected sentences per language, evaluated by two independent annotators. The results given in the table are the averages of the two independent evaluators' assessment.

Stanza model	LAS	UAS
EN EWT v1.2.3	85.94	89.08
DE GSD v1.2.3	85.78	91.03

Table 3.7: Parsing accuracy evaluation: observed for EPIC UdS V3.

By using measures to improve parsing accuracy such as taking out some of the spoken language features, correcting POS tagging for filled pauses, and using the smallest possible sentence segmentation, the observed parsing accuracy for EPIC UdS V3 is even higher than the stated parsing accuracy for the latest Stanza model. For a previous corpus version (EPIC UdS V2, described in Przybyl et al. (2022b)), using spacy NLP tools and including all the spoken features, the accuracy was ranging between 69.74 (LAS) and 86.42 (UAS). Even though V3 corpus data still include some spoken characteristics such as filled pauses, unfinished sentences, or no mid-sentence punctuation (which is especially relevant for detecting subordinate clauses, etc. in German), the results seem very reliable. Therefore, it can be assumed that the dependency relations are very accurate and are reliable input for the analysis in Chapter 6. However, it should be mentioned that interpreting data contain more spoken language features than original spoken language (Shlesinger and Ordan, 2012; Karakanta et al., 2021; Przybyl et al., 2022b), which might still have a minor effect on parsing accuracy. Accuracy for interpreting might be slightly lower than for original data. Furthermore, for long and incomplete sentences parsing accuracy is typically lower, which can be linked to originals (sentence length (Russo et al., 2012; Bernardini et al., 2016)) or interpreting (unfinished sentences). Nevertheless, the observed parsing accuracy for EPIC Uds V3 was assessed as very reliable overall.

3.2 Methodological design

As traditional corpus-based methods to investigate properties of interpreted language show mixed results (cf. Section 2.4), this thesis uses a two-level approach to investigate interpretese and the hypothesis of simplification in simultaneous interpreting. In a first step, a data-driven approach is used to detect typical features of interpretese (Chapter 4). Based on the results of this exploratory approach, simplification

as a potential characteristic of interpretese is investigated at two levels. Chapter 5 investigates lexical simplification and Chapter 6 focusses on syntactic simplification.

3.2.1 Exploring interpretese – Methods used in Chapter 4

The information-theoretic measure of relative entropy or Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951) has been proven to be suitable in corpus linguistics to show how one language variety differs from another language variety (Fankhauser et al., 2014; Degaetano-Ortlieb and Teich, 2018; Hughes et al., 2012; Klingenstein et al., 2014). Generally, it measures the number of additional bits that are needed to model a language variety when a non-optimal model is used. In this thesis, it is used to detect distinctive features of interpreted speech compared to original speech in the same language. This data driven approach is applied in Chapter 4 and aims to answer the general research questions on *interpretese* as laid out in Section 2.5.1:
RQ1a: Does interpreting differ significantly from original speech in the same language?
RQ1b: What are the linguistic feature that contribute to the specific linguistic properties of language that interpreters use?
If **RQ 1a** and the hypothesis that interpreting differs significantly from original speech in the same language is confirmed, **RQ1b** explores the specific properties that characterise *interpretese* in more detail by analysing the features that are found to be distinctive for one language variety compared to the other, using the method of relative entropy.

This exploratory approach uses strictly a comparable corpus design, comparing original spoken English to interpreted English from German as well as comparing original spoken German to interpreted German from English.

The advantages of using the method of relative entropy compared to more traditional frequency-based methods in corpus linguistics are as follows. It is a data-driven approach, in which no prior selection of features to be investigated is required. Furthermore, significance testing is built in, allowing for objective assessment and easier interpretation of results as only significant features are inherently considered. The approach used in this thesis is based on an implementation via word cloud visualisations (Fankhauser et al., 2014), which was also used in Karakanta et al. (2021). The method is described in more detail in Section 4.1 and applied in Section 4.2.

3.2.2 Exploring lexical simplification in interpreting – Methods used in Chapter 5

Some of the results of the exploratory approach on *interpretese* in Chapter 4 point to simplification as traditionally defined in corpus-based translation studies. These

results are further investigated in Chapter 5 as specific properties of interpreted language with a focus on lexical simplification due to the particular production constraints in interpreting, as laid out in Section 2.2. This chapter specifically addressed the following research questions:
RQ 2a: Is simultaneous interpreting more simplified at the lexical level than comparable originals in the same language?
RQ 2b: What are the particular triggers of lexical simplification in the source speech input? This question focusses on distinctive POS for interpreting which are mainly content words as the main carrier of a message.

First, selected traditional corpus-based measures as described in the theoretical background (Section 2.4.2) are used to approach a traditional definition of lexical simplification (cf. Section 2.5.2). The measures are selected based on the results of Chapter 4. The methods and results for traditional corpus-based measures on lexical simplification are given in Section 5.1.

As described in Section 2.4.2, traditional measures on simplification in simultaneous interpreting give unclear results. Research question **RQ 2a** first investigates interpreting via traditional corpus-based measures, investigating lexical density, token length, STTR and looks at POS distribution, and specific pattern length and variation. This thesis uses a more fine-grained approach to investigate traditional measures, considering source speech delivery type as well as source and target speech delivery rates, which could influence simplification tendencies. The approach uses a comparable approach, comparing interpreted speech to original speech in the same language (Section 5.1).

The main focus of the lexical investigation on simplification is the information-theoretic measure of *surprisal* or (un)expectedness in context to approach lexical simplification from a new angle. Surprisal is a measure that can be applied to the interpreting product (Rubino et al., 2016; Lapshinova-Koltunski et al., 2022; Kunilovskaya et al., 2023b) to study expected or unexpected sequences in the speech that would be considered more or less simple to comprehend for the audience, while providing a clear link to cognition (Hale, 2001; Teich et al., 2020) and the interpreting process (Carl, 2021; Wei, 2022; Lim et al., 2023).

The main hypothesis of this thesis is that the cognitively challenging conditions in simultaneous interpreting are believed to lead to simplification in the target product. For the lexical level, this is operationalised via more expected sequences in the interpreting product with lower surprisal. **RQ 2a** addresses this hypothesis further by investigating lexical simplification via the information-theoretic measure of surprisal, linking frequency in context to cognitive processing. This is investigated in a comparable corpus approach, comparing average surprisal in original and interpreting in the same language (English or German), for POS that were found typical for original or interpreted speech in Chapter 4. However, an expected tendency of interpreters to simplify may potentially be counterbalanced by properties of the source speech shining-through, such as the use of cognates (Chmiel et al., 2021) or potentially un-

expected translation solutions for unexpected source speech items. This is addressed in **RQ 2b**, looking into the triggers of lexical simplification via a parallel corpus approach, comparing source and target speech items (German into English and English into German).

The detailed methodology is described in Section 5.2.1, followed by the presentation of the analysis results in Sections 5.2.2 to 5.2.4.

3.2.3 Exploring syntactic simplification in interpreting – Methods used in Chapter 6

The results of the exploratory approach in Chapter 4 also point to syntactic differences in interpreting compared to original spoken language. Chapter 6 therefore takes a syntactic approach, using the dependency locality framework (Gibson, 1998, 2000) to operationalise syntactic simplification. In the same way that a tendency for optimal syntactic encoding can be observed in many natural languages (Futrell et al., 2015), interpreters may be particularly pressured to encode their output efficiently. Specifically, the following research questions are asked:
RQ 3a: Can syntactic simplification measured via sentence length (comparable and parallel approach) and alignment analysis (parallel approach) be observed in interpreting?
Furthermore, considering the aspects for and against potential dependency length minimisation, the following research questions are asked to give a more detailed insight about syntactic complexity.
RQ 3b: Can syntactic simplification measured via dependency length minimisation be observed in the interpreting product (comparable approach, comparing originals and interpreting in the same language)?
RQ 3c: What are the triggers of a potential minimisation effect? Does shining-through of source language structure counterbalance a possible minimisation effect (parallel approach, comparing source and target speech)?

Dependency locality measures can be used as a proxy of the sentence processing difficulty (Gibson, 2000; Liu, 2008) and applied to the interpreting product, less syntactic complexity is defined as more simplified (cf. Section 2.5.2). Before analysing different dependency locality measures, Section 6.2 first addresses **RQ 3a** and looks at overall sentence length in comparable and parallel interpreting data (Section 6.2.1) and takes a parallel approach by comparing source and target alignment of translation segments (Section 6.2.2). Although previously discussed in the literature that interpreters use shorter sentences than original speakers (Bernardini et al., 2016; Liu et al., 2023) and interpreters possibly split source sentences (Kalina, 1998; He et al., 2016; Gast and Borges, 2023), this thesis is the first study to our knowledge that investigates large-scale source-to-target aligned interpreting data to study how interpreters deal with source sentences in their target output. After assessing the source-to-target sentence alignment, the main focus is then on **RQ 3b**, investigating syntactic simpli-

fication in simultaneous interpreting via dependency length measures such as average dependency length, maximum dependency length, or number of adjacent dependencies within a sentence in a comparable corpus approach, comparing originals and interpreting in the same language (English and German). As with other measures of simplification in interpreting, the pressure to optimise encoding of the target speech may be counterbalanced by source speech and language properties. This is addressed in **RQ 3c**, looking into the triggers of a potential minimisation effect, using a parallel corpus approach, comparing source and target speech (German into English and English into German).

Section 6.1 describes the methodology applied for the syntactic study, followed by a presentation of the results of the syntactic analysis in Section 6.2.

Part III

Investigations

Chapter 4

Exploring general properties of interpretese

Translations differ from original production as discussed in Section 2.1. To explore if interpreted language also differs from original spoken production and find the features that characterise interpretese, the information-theoretic method of relative entropy or KLD (Kullback and Leibler, 1951) is used in this Chapter. Relative entropy is used to calculate the difference between the distribution of interpreting and original spoken production, and aims to answer the research questions as laid out in Section 2.5.1:
RQ 1a: Does interpreting differ significantly from original speech in the same language?
RQ 1b: What are the linguistic features that contribute to the specific linguistic properties of language that interpreters use?

First, the method of relative entropy is explained in Section 4.1. The results are then discussed in Section 4.2, followed by an evaluation of the research questions 4.3. Last, individual results of the exploratory analysis are selected in the context of the research questions on simplification and are studied in more detail in Chapter 5 (lexical investigations) and Chapter 6 (syntactic investigations).

4.1 Exploratory approach to interpretese via relative entropy: method

Relative entropy can be used to detect distinctive features of interpreted speech compared to original speech in the same language(s). The method is used in Section 4.2 to detect similarity/differences at the lexical level, visualised via word clouds (Fankhauser et al., 2014). The method of relative entropy has been shown to be suitable in corpus linguistics to detect linguistic varieties and styles (Degaetano-Ortlieb and Teich, 2018; Hughes et al., 2012; Klingenstein et al., 2014). Fankhauser et al.

(2014) apply relative entropy to detect register differences in the English Brown Corpus, whereas Degaetano-Ortlieb and Teich (2018) show its application in diachronic change. The method used in this thesis is based on Karakanta et al. (2021) and the results are published in Przybyl et al. (2022a).

The general approach is as follows. Probabilistic language models (LMs) (here: unigram LMs) are built for originals and interpreted speeches for German and English individually (cf. corpus data in Section 3.1, using words as features for the lexical investigations). The obtained LMs are then used to calculate the relative entropy (KLD) between the models by estimating the amount of additional information bits that are needed to model originals by interpreted speech and vice versa (Equation 4.1). The results are the linguistic features (here: words) that contribute most to the difference, i.e. features that are most typical for one distribution compared to the other (cf. Karakanta et al. (2021) for a more detailed description).

$$D(A||B) = \sum_i p(feature_i|A) log_2 \frac{p(feature_i|A)}{p(feature_i|B)} \tag{4.1}$$

Equations 4.2 and 4.3 show the two general approaches taken in the thesis. Original speeches (e.g. ORG EN or ORG DE) are compared to interpreted speeches (e.g. SI DE EN or SI EN DE) and vice versa. As KLD is an asymmetric measure, the list of distinctive features resulting from Equation 4.2 may differ compared to the results of 4.3. Equation 4.2 measures the difference between SI and ORG with SI used as a basis for encoding. Equation 4.3 measures the difference between ORG and SI with ORG used as a basis for encoding.

$$D(SI||ORG) = \sum_i p(feature_i|SI) log_2 \frac{p(feature_i|SI)}{p(feature_i|ORG)} \tag{4.2}$$

$$D(ORG||SI) = \sum_i p(feature_i|ORG) log_2 \frac{p(feature_i|ORG)}{p(feature_i|SI)} \tag{4.3}$$

For the lexical investigations in Section 4.2 a word-based unigram KLD model is used to detect lexical features that are distinctive for original and interpreted speech. Jelinek-Mercer smoothing for different types of text was used on the LMs. Preprocessing for the word-based unigram model included tokenisation, lowercasing, and number conflation. Post-processing of the results included the manual exclusion of truncated words from the results (cf. Karakanta et al. (2021)).

KLD scores of individual words are visualised using word clouds (Fankhauser et al., 2014). An unpaired Welch t-test on the observed word probabilities for each speech in the corpus is used to assess statistical significance. Only words with a $p-value \leq 0.05$ are displayed in the word cloud and are evaluated in the analysis. In the word clouds, distinctivitiy is visualised by size (the bigger the word, the more distinctive it is within the distribution). The colour represents the relative frequency (red = high frequency; blue = low frequency).

4.2 Interpretese detection via relative entropy

To detect distinctive characteristics of *interpretese*, the distances between the LMs for German originals vs. German interpreting (from English as a source) and vice versa are investigated, as well as the same comparison in English. The distinctive features obtained are visualised as word clouds (Figures 4.1 and 4.2) and listed as Tables 4.1 to 4.4. The features obtained differ depending on whether the LM for interpreting is compared to the LM for originals or the distance between originals compared to interpreting is calculated. The results are also described by Przybyl et al. (2022a).

(a) Interpreting (b) Spoken originals

Figure 4.1: German simultaneous interpreting (a) vs. spoken originals (b). Distinctivity is visualised by size, frequency by colour.

The results for German and English show the same tendencies overall. Simultaneous interpreting (SI) in both languages is characterised by more spoken language features than spoken originals (Figure 4.2 and 4.1), (a) Interpreting): Hesitations (*euh, hm, hum*) are by far the most distinctive word unigramms for SI. They are visualised as highly frequent (red) and very distinctive (large font) in the word cloud, with distinctivity values of 0.03960763 (for EN *euh*), 0.00698372 (for EN *hum*) and 0.13025398 (for DE *euh*) (cf. Tables 4.1 and 4.2). This means that, e.g. 0.13025398 extra bits of information would be needed in order to encode *hum* in the German original spoken distribution, based on the distribution of German simultaneous interpreting[1].

Intensifiers are another typical spoken feature that is distinctive in the results and therefore characteristic for SI. For German interpreting, this includes *wirklich*, *ganz*[2], and *so*. For English interpreting, the intensifiers *really* and *obviously* are

[1] A list of all KLD values, frequencies and p-values used as a basis for the word clouds is given in the Appendix: Tables A to A.4.

[2] The word based KLD approach does not take into account grammatical function. Therefore, *so* could also be used as modal adverb or *ganz* as adjective.

word	relative entropy	frequency	p-value
euh	0.13025398	0.06117160	0.00000000
also	0.00991705	0.00419329	0.00000001
hum	0.00805377	0.00383335	0.00002856
hm	0.00674845	0.00185368	0.00000013
wirklich	0.00542696	0.00338342	0.00013821
ja	0.00418684	0.00478719	0.00029221
aber	0.00404127	0.00689283	0.00138020
hier	0.00404115	0.00631693	0.00704150
jetzt	0.00400544	0.00503914	0.00013157
haben	0.00378935	0.01018627	0.00760990
dann	0.00346510	0.00453523	0.00154950
möchte	0.00288743	0.00311347	0.00583070
müssen	0.00252322	0.00464321	0.00819820
ganz	0.00221240	0.00233960	0.00248550
denn	0.00212316	0.00232161	0.00066488
so	0.00179292	0.00372537	0.02946700
sicherstellen	0.00175733	0.00034194	0.00005351
arbeiten	0.00168127	0.00077387	0.00182150
möchten	0.00164440	0.00050391	0.00301260
war	0.00135383	0.00196167	0.01476700
gab	0.00119398	0.00059390	0.03523400
sagen	0.00117587	0.00228561	0.00483700
freue	0.00097261	0.00034194	0.02118500

Table 4.1: KLD values as basis for word cloud visualisation: SI EN DE vs. ORG DE.

distinctive. SI is also more distinctive with regard to the use of verbs, including modals, auxiliaries, and very general lexical verbs. For German interpreting, *haben*, *möchten*, *möchte*, *war*, *müssen*, *sicherstellen*, *arbeiten*, *sagen*, *gab*, and *freue* are found distinctive when comparing the German interpreting LM to the LM for German originals. For English interpreting, distinctive verbs are *be*, *need*, *can*, *going*, *talking*, *shouldn*, and *react*[3].

Taking the perspective of originals (ORG) and modelling them against interpreting gives a completely different result, again, similar in both languages investigated (Figure 4.2 and 4.1, (b) Original spoken). The most distinctive item for German originals compared to interpreting is *der* with a KLD value of 0.01547948 (cf. Tables 4.3

[3] *gambling* is not included in the list of verbs as a context analysis revealed only nominal use of this lexical item in the corpus.

word	relative entropy	frequency	p-value
euh	0.03960763	0.04218045	0.00000000
hum	0.00698372	0.00375940	0.00001041
we	0.00596984	0.02483083	0.02644400
be	0.00460203	0.01135338	0.00547490
't	0.00317923	0.00535714	0.01024100
need	0.00315935	0.00411654	0.00293900
're	0.00312006	0.00413534	0.00359710
can	0.00306121	0.00518797	0.00601240
going	0.00264991	0.00225564	0.00377160
talking	0.00238054	0.00131579	0.00445210
really	0.00154350	0.00184211	0.02499000
obviously	0.00120633	0.00054511	0.00863260
shouldn	0.00098056	0.00050752	0.02400800
'd	0.00073079	0.00129699	0.02472500
react	0.00058932	0.00011278	0.03180000

Table 4.2: KLD values as basis for word cloud visualisation: SI DE EN vs. ORG EN.

(a) Interpreting (b) Spoken originals

Figure 4.2: English simultaneous interpreting (a) vs. spoken originals (b). Distinctivity is visualised by size, frequency by colour.

and 4.4 for distinctive features for originals). Although the lexeme can also be used as a relative pronoun, the list of other distinctive elements contains other determiners and therefore points to a nominal use in ORG compared to SI. The same trend can be observed in both languages: DE: articles *des*, *den*[4], indefinite pronouns *einer*,

[4] *den* may also be used as relative pronoun.

keine, possessive pronouns *meine*, *unserer*, and EN: articles *an*[5], possessive pronouns *our*, *my*, *their*, *his*, and demonstrative pronouns *those*. Furthermore, nouns can be detected as distinctive for original production. This may be a result of topical differences in the texts of both distributions. However, in combination with the distinctive use of determiners in ORG, original spoken for English and German clearly shows more nominal features.

word	relative entropy	frequency	p-value
der	0.01547948	0.03094852	0.00000001
in	0.00645700	0.01761912	0.00378820
des	0.00388808	0.00592827	0.00033493
mit	0.00384489	0.00771412	0.00162740
von	0.00366855	0.00872671	0.00252830
als	0.00288532	0.00430812	0.00616070
den	0.00240571	0.01056779	0.02214200
meine	0.00202994	0.00156492	0.00076597
nach	0.00186599	0.00215406	0.00568870
einer	0.00184265	0.00244863	0.00649120
aus	0.00176983	0.00298255	0.01313700
zur	0.00129158	0.00182267	0.02514700
unserer	0.00128455	0.00092054	0.00527330
keine	0.00112995	0.00156492	0.04605400
am	0.00084152	0.00136240	0.03074400

Table 4.3: KLD values as basis for word cloud visualisation: ORG DE vs. SI EN DE.

Furthermore, distinctive features for ORG are multiple prepositions as visualised by *in*, *mit*, *von*, *als*, *nach*, *aus*, *zur*, and *am* for German and *on*, *by*, as well as *at* for English. This might point towards longer complex sentences in originals as compared to interpreting.

From a more general perspective, it can be observed that fewer and larger items are distinctive for SI, which are also used with higher frequencies (red, orange, and yellow colour for SI; more blue, and green items in ORG). This points to less lexical variation in simultaneous interpreting than in spoken originals and, therefore, more repetition in vocabulary used by interpreters.

Some language-specific differences can also be observed. Fewer words are distinctive in German than in English (fewer items that contribute significantly to the

[5]Note that the item with the second highest KLD value for EN ORG is also *the*, KLD-value: 0.00627184, but the p-value of 0.07359200 shows that it is not significant and therefore not included in the word cloud.

word	relative entropy	frequency	p-value
i	0.00691985	0.01811679	0.00690470
on	0.00470029	0.00987273	0.00390810
our	0.00330152	0.00505390	0.00909200
my	0.00325766	0.00256892	0.00017945
our	0.00330152	0.00505390	0.00909200
my	0.00325766	0.00256892	0.00017945
by	0.00238901	0.00359314	0.00645790
their	0.00181168	0.00315659	0.01448400
his	0.00155837	0.00104100	0.01494700
at	0.00147261	0.00493636	0.03092100
those	0.00125006	0.00230028	0.03952600
me	0.00120356	0.00112495	0.03329900
an	0.00103109	0.00357635	0.04809700

Table 4.4: KLD values as basis for word cloud visualisation: ORG EN vs. DI DE EN.

difference between the two distributions). This points towards less variation in German overall. German interpreting is characterised by additional spoken features when compared to German originals: discourse markers/particles (*also*, *ja*), deictics (*hier*, *jetzt*), but also conjunctions/conjunctive adverbs (coordinating *aber*, *denn*, and subordinating *dann*) are shown in the interpreting word cloud, again pointing to syntactic differences.

For English, other typical spoken elements that are distinctive for SI and not ORG are contractions (*'re*, *'t*, and *'d*,). We can observe English originals referring more to the speaker themselves (*I*, *our*, *my*) while interpreters opt for *we*. In addition to the differences of SI being more verbal than originals, in English we can also see differences in aspect (progressive form or gerund clauses of *going*, and *talking*). The word cloud inspection does not give information on the use of the word itself. Further analysis would be required to identify gerund clauses or verb tense.

4.3 Summary of interpretese detected via relative entropy

To summarise the exploratory analysis using relative entropy to detect *interpretese*, it can be stated that the distributions of interpreting and original production for both languages studied differ significantly. This is shown via specific items characterising interpreting but not originals and vice versa and confirms **RQ 1a**: Interpreting differs significantly from original production in the same language. As for **RQ 1b**, it can

be stated that the most obvious difference is that there are more spoken language features in simultaneous interpreting than in spoken originals in the same language. Such a trend was already observed by Shlesinger and Ordan (2012) who described simultaneous interpreting when compared to translations to be characterised rather by the spoken mode than translation features. The present study confirms such a trend also for interpreting when compared to spoken originals, particularly showing that this relates to filled pauses, intensifiers, deictics, discourse makers, and more verbal use. Compared to interpreting, originals are found to be more nominal, including distinctive nouns and determiners, but also using different syntactic structures, as indicated by more distinctive prepositions. These observations are true for both languages studied here, the comparable English dataset as well as comparable German interpreting and originals. Similar results with more nominal structures and complex main clauses in originals have also been shown for English and Spanish (Karakanta et al., 2021) using the same method. Furthermore, Gast and Borges (2023) also observe differences in POS distributions for interpreting and originals related to verbal and nominal use, pointing to syntactic differences in both subsets. As some structures such as complex nominals are regarded as more complex than others (Lu, 2010), more verbal use in interpreting relate to simplification of the interpreting product.

Regarding other indicators of simplification, less lexical variation in interpreting, as found in the literature (Russo et al., 2012; Kajzer-Wietrzny and Ivaska, 2020; Xu and Li, 2022), can also be observed via the method applied in this study.

The differences observed in this study as specific properties of interpreted language or *interpretese* may be a result of specific production constraints dealt with by the interpreter. However, it also has to be taken into account that some of the original speeches in the European Parliament (and part of this corpus) are prepared speeches that are read out by the original speakers. Interpreting, on the other hand, can be seen as spontaneous delivery. Further analysis using information on delivery type of the speeches (originals being prepared and read out, impromptu speeches, or mix deliveries) is necessary to draw further conclusions on the origin of this difference in spokenness.

The results of the KLD analysis are used for further investigations on the lexical level (Chapter 5) and the syntactic level (Chapter 6), and can be linked as follows. For the lexical level, some of the observed differences between originals and interpreting can be related to traditional corpus-based translationese measures (Section 5.1). Differences in the distribution of function words and lexical words are measured using lexical density. For differences in the length of words that are typical for each subset, token length and pattern length are investigated further. As lexical variation was observed to be lower in interpreting, detected with relative entropy, the measure of (standardised) type-token ratio is applied in order to confirm or reject the result with a traditional measure.

Furthermore, it was observed that interpreting seems to distinctively use more, but also more general or high frequent verbs and nouns. In order to study whether there

is a difference in expectedness of verbs and nouns (as there are among the distinctive POS categories with respect to relative entropy), the information-theoretic measure of surprisal is applied to further investigate the differences (Section 5.2.2).

Verbal vs. nominal POS in interpreting when compared to originals, originals using more distinctive prepositions, while interpreting uses conjunctions point to syntactic differences. A theory of syntactic complexity that takes into account optimal encoding from a memory perspective is DLT. Syntactic complexity is therefore investigated in Chapter 6.

Chapter 5

Exploring lexical simplification

The results of the exploratory analysis via relative entropy point to differences in the use of lexical words when comparing interpreting to original speech in the same language. This chapter addresses the research questions on lexical simplification, specifically laid out as:
RQ 2a: Is simultaneous interpreting more simplified at the lexical level than comparable originals in the same language?
RQ 2b:What are the particular triggers of lexical simplification in the source speech input? This question focusses on distinctive POS for interpreting which are mainly content words as the main carrier of a message.
Differences in the use of individual POS (e.g. use of verbs, nouns, prepositions) can be studies with standard measures of simplification such as lexical density, token length, and STTR. **RQ 2a** on lexical simplification is first investigated via traditional corpus-based measures in Section 5.1. Second, **RQ 2a** and **RQ 2b** on lexical simplification studied via the information-theoretic measure of surprisal and the trigger of distinctive differences are investigated in Section 5.2.2, first taking a comparable approach to study the general tendency of interpreters to simplify (**RQ 2a**) and then a parallel perspective to look at individual occurrences of simplification (**RQ 2b**).

Lexical simplification for this purpose has been defined as interpreted or original speech that uses lower variation, lower lexical sophistication, and lower level of information content. This particularly includes more expected words in context, i.e. a high degree of lexical simplification is indicated by the use of low surprisal words on average, whereas unexpected words in context can be linked to a low degree of simplification (cf. Section 2.5.2).

5.1 Selected traditional corpus-based measures of simplification

This section investigates traditional measures of simplification such as lexical density (LD), token length and standardised type-token ratio STTR and patterns in a comparable approach, comparing interpreting and originals in the same language. However, instead of using a traditional approach and investigating general simplification measures, the focus of this approach is to investigate corpus-based measures of simplification that are derived from the exploratory analysis (cf. Section 4). The analysis also includes additional parameters that might influence simplification tendencies, such as source speech delivery type, or source and target speech delivery rate. The basic principle used here is Blum-Kulka and Leverston's definition of lexical simplification, which is "the process and/or result of making do with less words" (Blum and Levenston, 1978, p.399). In this sense, Laviosa (1998) sees lexical simplification as the tendency to greater repetitiveness and, therefore, lower level of information content and lower lexical sophistication.

5.1.1 Lexical density

A major result of investigating the distinctive features for interpreting vs. original speech (Chapter 4) is a difference in the use of various POS. Nominal word classes are found to be more distinctive for original speech, whereas verbs are seen as distinctive for interpreting. A standard measure to account for this contrast is lexical density (LD), calculated as the number of lexical words divided by the total number of words in the corpus. A speech would be defined as more simplified if it contains more functions words and fewer lexical words (thus a lower lexical density). Therefore, a more simplified text or speech would be defined as a text or speech with more function words and fewer lexical words. Lower lexical density is defined as more simplified.

$$LD = \frac{lexical\ words}{total\ words} * 100 \tag{5.1}$$

Lexical density calculated in this section is based on the universal POS tags in the EPIC-UdS V3 corpus version. The POS classes used as lexical words are noun (NOUN), proper noun (PROPN), verb (VERB), adjective (ADJ) and adverb (ADV). The result slightly differs from lexical density stated in Przybyl et al. (2022a), which was based on a previous corpus version, while this corpus version uses more reliable POS tagging, as some spoken features were excluded before tagging. Table 5.1 gives lexical density for the different subcorpora.

Original speeches in both languages are significantly lexically denser[1] and therefore interpreting can be seen as simpler than comparable originals for this indicator of

[1]Pearson's Chi-squared test with Yates' continuity correction - EN: X-squared = 9.6273, p-value = 0.001917; DE: X-squared = 67.623, p-value < 2.2e-16

English	ORG SP EN	SI DE EN
LD	44.39	43.02
German	ORG SP DE	SI EN DE
LD	47.43	43.57

Table 5.1: Lexical density (LD) for EPIC UdS V3.

simplification. As the analysis of relative entropy showed a tendency of interpreting using more spoken features than spoken originals, a more fine-grained analysis of source speech delivery types is carried out. The source speeches in the corpus include speeches that were prepared and read-out by MEPs (read), as well as freely delivered speeches (impromptu). Whenever the speaker was holding a script, however, did not look at the notes most of the time, the delivery type was categorised as mixed (cf. Section 3.1). Read-out speeches are expected to show more traces of written language. We therefore compare only impromptu speeches to interpreted speeches, which are defined as spontaneous production for this corpus[2]. Table 5.2 gives lexical density for the different source text delivery types and interpreted speech.

English	ORG SP EN	ORG SP EN	ORG SP EN	SI DE EN
	read	mixed	impromptu	(impromptu)
LD	48.23	43.99	42.47	43.02
German	ORG SP DE	ORG SP DE	ORG SP DE	SI EN DE
	read	mixed	impromptu	(impromptu)
LD	49.61	46.42	46.22	43.57

Table 5.2: Lexical density (LD) with respect to delivery type for EPIC UdS.

The trend observed for lexical density on the whole dataset (interpreting less dense than comparable originals) cannot be confirmed when considering the preparedness of the original speeches. As expected, prepared and read-out source speeches are significantly denser than mixed and impromptu speeches[3]. When comparing the two spontaneous speech production deliveries in English (impromptu original speeches and interpreting), we cannot see significant differences for impromptu originals vs. simultaneous interpreting[4]. For German, the trend of interpreted speeches being simpler than any of the original speech delivery types can be confirmed: lexical density in read speeches is lower than in the mixed production type. Impromptu speeches show even lower lexical density with simultaneous interpreting having lowest lexical

[2]Due to the short length of the interventions and the large number of speakers in a plenary session, we expect interpreters not to be able to produce an interpretation with text.

[3]Pearson's Chi-squared test with Yates' continuity correction - EN: X-squared = 47.731, p-value = 2.43e-10; DE: X-squared = 88.432, p-value < 2.2e-16

[4]Pearson's Chi-squared test with Yates' continuity correction X-squared = 0.6183, p-value = 0.4317

density[5].

English	SI DE EN	SI DE EN	SI DE EN
	from read source	from mixed source	from impromptu source
LD	43.50*	(43.20)	42.32*
German	SI EN DE	SI EN DE	SI EN DE
	from read source	from mixed source	from impromptu source
LD	44.47*	(43.22)	43.73*

Table 5.3: Lexical density (LD) for SI in EPIC UdS with respect to source delivery type.

In the next step, additional parameters that might influence lexical density are included in the analysis in order to exclude other possible factors influencing the observed lexical simplification as discussed above. Lexically denser input of the source might affect lexical density in interpreter's output. First, delivery types of the source speeches on which the interpretations are based are taken into account. Table 5.3 focusses on lexical density in interpreted speech with respect to this variable. Although interpreting based on read-out sources is slightly denser than interpreting based on impromptu source speeches, these differences are not significant, neither in English, nor in German[6].

A further factor that potentially could influence lexical density in original and interpreted speeches is the rate of delivery of source and target. This is encoded in the corpus metadata as delivery rates in words per minute (wpm), grouped in rates of *high* ($\geq$ 161 wmp), *medium* (131-160 wpm) and *slow* (*slow* $\leq$130 wpm) (cf. Section 3.1). Original and interpreted speeches of the same delivery rate (e.g. delivery rate *high* ORG SP EN vs. SI DE EN) are compared to confirm the above stated trend of interpreted speech being less lexically dense than originals. Lexical density as shown in Table 5.4 shows lower values for interpreted speech vs. originals of the same delivery rate (e.g. delivery rate high ORG SP EN 43.72 vs. SI DE EN 42.51). This holds for both languages and across all delivery rates, however, is only significant for English delivery rate medium and German delivery rate high and medium[7]. The results on delivery rate confirm the overall observation that SI is

[5]Pearson's Chi-squared test with Yates' continuity correction X-squared = 15.921, p-value = 6.605e-05

[6]Pearson's Chi-squared test with Yates' continuity correction - *EN: X-squared = 2.466 - not significant, p-value = 0.2914; *DE: X-squared = 2.1748, p-value = 0.3371 - not significant

[7]Pearson's Chi-squared test with Yates' continuity correction - *EN high: X-squared = 2.9872, p-value = 0.08393 - not significant; EN medium: X-squared = 17.186, p-value = 3.39e-05; *EN slow: X-squared = 1.7245, p-value = 0.1891 - not significant; DE high: X-squared = 11.853, p-value = 0.0005757; DE medium: X-squared = 48.614, p-value = 3.116e-12; *DE slow: X-squared = 1.7245, p-value = 0.1891 - not significant

lexically simpler than ORG.

English	ORG SP EN			SI DE EN		
	high	medium	slow	high	medium	slow
LD	43.72*	46.38	45.85*	42.51*	43.10	43.51*
German	ORG SP DE			SI EN DE		
	high	medium	slow	high	medium	slow
LD	46.62	47.68	47.52*	43.09	43.44	44.86*

Table 5.4: Lexical density (LD) for EPIC UdS with respect to delivery rate.

To summarise the results on lexical density for the comparable English and German dataset, it can be stated that generally a trend of simplification in interpreting can be observed, indicated by lower lexical density in interpreting than comparable originals. This trend is reversed than the densification trend for interpreted English stated in the literature (Kajzer-Wietrzny, 2015; Sandrelli and Bendazzoli, 2005; Ferraresi et al., 2018; Lv and Liang, 2019; Dayter, 2018). However, this trend was also not confirmed across all source languages or language varieties in the literature (Ferraresi et al., 2018; Liu and Dou, 2023). The results of this section show lexical density in the context of further source and target speech variables. When comparing impromptu originals and interpreting, lower lexical density can be found for German interpreting, while no significant differences are observed for comparable English impromptu speeches. When comparing speeches of the same delivery rate, interpreting can generally be found to be less lexically dense than comparable originals. Furthermore, source speech delivery type alone does not influence lexical density in interpreting.

5.1.2 Token length

The method of relative entropy makes characteristic items of a distribution visible and shows that distinctive words in interpreting seem generally shorter than words distinctively used in originals (Chapter 4). This relates to word length as a simple measure of lexical sophistication. Mean token length is therefore reported here as an indicator of simplicity of token used, relating shorter words to more simplicity. Table 5.5 shows the results for token length. Mean token length is significantly shorter in interpreting than originals for both languages[8]. This is also confirmed by shorter median token length in German interpreting compared to German originals (Table 5.5).

As we observe a difference in use of function vs. lexical words via the method of relative entropy, we also specifically look at token length of lexical words. Mean and median token length of the lexical word classes (NOUN, PROPN, VERB, ADJ and

[8]Welch two sample t-test EN: t = 2.1541, df = 124164, p-value = 0.03123; DE: t = 17.902, df = 112935, p-value < 2.2e-16

English	ORG SP EN	SI DE EN
mean token length	4.50	4.47
median token length	4.0	4.0
German	ORG SP DE	SI EN DE
mean token length	5.94	5.55
median token length	5.0	4.0

Table 5.5: Mean token length.

ADV) is calculated. The results are given in Table 5.6. The same trend as for general token length is observed. Interpreting shows a tendency to use shorter lexical words than comparable originals. Mean token length of lexical words is statistically shorter for interpreting in both languages, compared to originals in the same language, with a stronger effect again in German than English[9]. Median token length for lexical words does not show a difference.

English	ORG SP EN	SI DE EN
mean token length	5.84	5.78
median token length	5	5
German	ORG SP DE	SI EN DE
mean token length	8.16	7.71
median token length	7	7

Table 5.6: Mean token length for lexical words.

The measure of mean token length relating to lexical sophistication – for all tokens and for lexical words – points to simplification in interpreting compared to originals in the same language.

5.1.3 Standardised type-token ratio

A further measure to give an overview of simplicity or variation of a dataset is type-token ratio. As the individual subcorpora vary slightly in size, standardised type-token ratio (STTR) is used to give an indication of lexical variation. Table 5.7 shows less lexical variation and therefore more simplification for interpreting than for originals in both languages.

This is in line with the literature on this measure for original English compared to interpreted English based on various source languages (Liu and Dou, 2023; Lv and Liang, 2019; Ferraresi et al., 2018). Interpreted English can be stated to be less repetitive and therefore more simplified than original English. The result given

[9]Welch two sample t-test EN: t = 3.3896, df = 68465, p-value = 0.0007002; DE: t = 13.681, df = 57310, p-value < 2.2e-16

English	ORG SP EN	SI DE EN
STTR	0.40	0.38
German	ORG SP DE	SI EN DE
STTR	0.47	0.43

Table 5.7: Standardised type-token ratio.

here also extends this trend found in the literature for English to interpreted German compared to German originals.

5.1.4 POS distribution

The results of the exploratory analysis using relative entropy show distinctive differences in the use of certain POS. Interpreting is characterised by words belonging to the POS classes adverb (ADV), verb (VERB), auxiliary (AUX), conjunctions (subordinating conjunction (SCONJ) and coordinating conjunction (CCONJ)) and interjection (INTJ), whereas originals are characterised by nominal word classes (noun, (NOUN), proper noun (PROPN), pronoun (PRON) and determiner (DET)) as well as adposition (ADP). Therefore, a POS distribution of these POS is carried out to show whether these differences only relate to individual words or are rather more general differences.

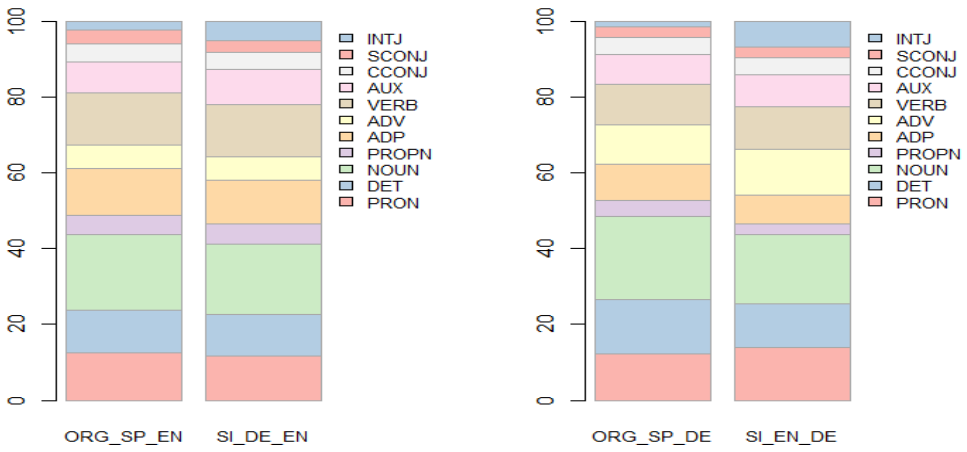


Figure 5.1: POS distribution.

Figure 5.1 shows that not only the individual words belonging to the POS classes, shown in the word cloud visualisation, are distinctive for a subset, but for some POS it is the whole word class that is used more frequently in a subset. As expected, comparing POS distributions of originals vs. interpreted shows distinctive differences: interjections (mainly due to filled pauses *euh, hm* and *euh*) and auxiliary verbs are more frequently used in interpreted speech; nominal word classes (determiners, nouns,

and proper nouns) are more frequently used in originals (cf. Table 5.8)[10].

subcorpus	fpm VERB	fpm AUX	fpm NOUN	fpm DET	fpm INTJ
ORG EN	116,242.22	68,031.23	168,500.81	95,889.03	18,855.15
SI DE EN	113,997.22	77,128.79	154,551.09	89,505.27	40,912.50
ORG DE	91,659.51	65,497.55	186,519.77	120,685.30	10,815.18
SI EN DE	95,275.80	72,232.12	154,905.65	100,438.77	57,172.08

Table 5.8: Frequency per million for selected POS.

The POS distribution confirms the results of the exploratory analysis for distinctive words, with more frequent use of verbal POS in interpreting and nominal POS in originals. Similar observations are made in the literature (Gast and Borges, 2023). Therefore, nominal and verbal POS are investigated in more detail in Sections 5.2.3 and 5.2.4.

5.1.5 Pattern length and variation: noun phrases

The results of the exploratory analysis point to a more nominal style in original speech. Lexical words in general are found to be shorter in interpreting than originals (cf. Section 5.1.2). As speakers have different possibilities to encode nominal referents varying in length, distribution of noun phrases are compared. This includes the shortest encoding of using a pronoun (PRON), as well as longer noun phrases up to determiner-adjective-adjective-noun (DET-ADJ-ADJ-NOUN). The results are shown in Figure 5.2.

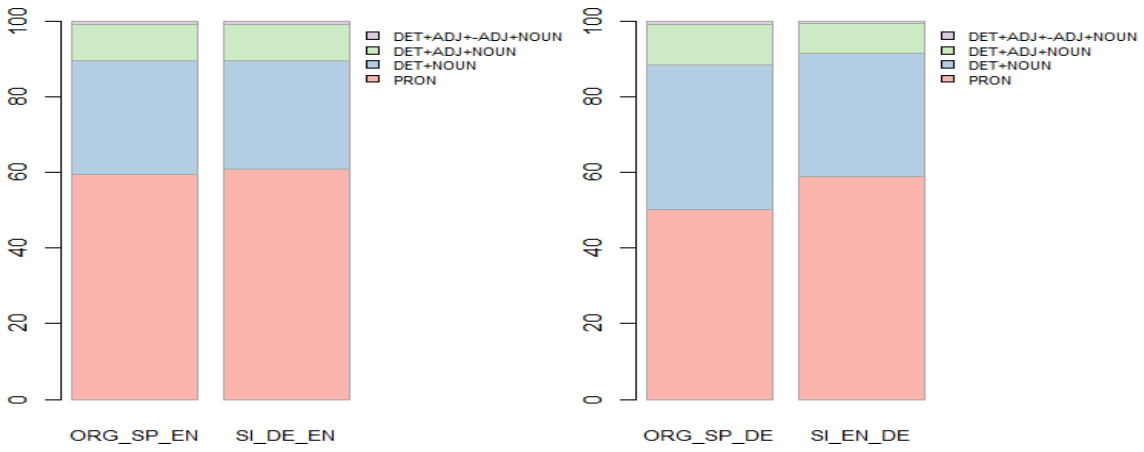


Figure 5.2: Noun phrase distribution.

[10]Pearson's Chi-squared test - EN: X-squared = 693.84, p-value < 2.2e-16;DE: X-squared = 2561.8, p-value < 2.2e-16

The noun phrase distribution for English only shows a slight tendency of shorter patterns used in interpreting, which does not prove to be significant[11]. For German, on the other hand, there is a clear preference in interpreting to use shorter encodings more frequently than longer noun phrases[12]. In particular, the patterns DET-ADJ-NOUN and DET-ADJ-ADJ-NOUN are used less frequently in German interpreting than in original speech. German interpreting prefers the use of pronouns. Whether this preference for shorter encodings is purely due to time constraints or potentially also due to omission of modifiers (cf. Section 2.2.2), cannot be answered here. Furthermore, shorter encodings could potentially carry more information. Chapter 5.2 further investigates the information content of the linguistic units used by interpreters and original speakers.

5.1.6 Summary of traditional corpus-based measures of simplification

Focussing on distinctive differences between original spoken language and interpreting (results of Chapter 4) when using corpus-based measures of simplification gives clearer results than the general approaches to simplification taken in the literature (Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2015; Ferraresi et al., 2018; Dayter, 2018; Kajzer-Wietrzny and Ivaska, 2020; Lv and Liang, 2019). For the language pair German-English, results discussed in this section clearly show signs of simplification in simultaneous interpreting, indicated by lower lexical density, not just overall, but also when focussing on impromptu delivery and speeches at the same delivery rate. Simplification trends observed here also include lower lexical sophistication and lower lexical variation. Interpreting generally uses shorter encodings (word length as well as noun phrases). The simplification trends observed are more pronounced for German interpreting compared to German originals, but also clearly visible for the English comparable dataset. **RQ 2a** on lexical simplification in simultaneous interpreting compared to original speech in the same language can be confirmed for the language pair German and English using traditional measures.

5.2 Investigating lexical simplification via surprisal

The remainder of this chapter addresses lexical simplification from a new angle. It investigates the question whether interpreters simplify their output by using more expected words in context, operationalised via the information-theoretic method of *surprisal*. Via a comparable approach, comparing original and interpreting in the same language, **RQ 2a** asking whether simultaneous interpreting is more simplified at the lexical level than comparable originals in the same language is addressed. In

[11]Pearson's Chi-squared test - *EN: X-squared = 5.6856, p-value = 0.1279 - not significant
[12]Pearson's Chi-squared test - DE: X-squared = 200.56, p-value < 2.2e-16

a parallel approach, comparing source and target speeches, **RQ 2b** is addressed by focussing on the particular triggers of lexical simplification. In particular, distinctive POS for interpreting, mainly content words as the main carrier of a message, are investigated. Section 5.2.1 describes the method used in the analysis and Sections 5.2.2, 5.2.3 and 5.2.4 analyse the results. Section 5.2.5 summarises the findings of the approach to lexical simplification investigated via surprisal.

5.2.1 Surprisal as measure of simplification: method

Texts are assumed to vary in the level of information content they convey. The information content encoded in a linguistic unit, e.g. a word, can be measured as *surprisal* (Hale, 2001). Surprisal will be high when a word in it's context has a low probability. On the other hand, words with low surprisal are highly expected in their context. Information that texts convey is assumed to be distributed evenly over a message, aiming to avoid information peaks as well as low informative points (Levy and Jaeger, 2006). Informativity above or below expected rates may lead to processing difficulties (Engelhardt et al., 2011; Kravtchenko and Demberg, 2015). As surprisal for individual words correlate with measured reading times (Levy, 2008; Smith and Levy, 2013; Bicknell and Levy, 2010), it can be seen as measure of processing complexity. Higher word surprisal leads to higher integration costs, which accounts for prolonged gaze duration (Demberg and Keller, 2008). Degaetano-Ortlieb and Teich (2022) assume that linguistic options available to the encoder (in our case the original speaker or interpreter) are used to modulate the level of information encoded in the message to achieve optimal encoding. This approach takes a production perspective.

Surprisal as an information-theoretical approach to formalise *unpredictability* in context is used as follows. Surprisal is measured as information content in number of bits, calculated as negative log base 2 probability of a given occurrence of a unit (e.g. word) in a given context (cf. Equation 5.2). Context is here defined as a number of preceding words.

$$S(word) = -log_2 p(word|context) \tag{5.2}$$

Surprisal for each word is annotated in corpus version EPIC-UdS (aligned, parsed, surprisal). The language models used as basis for calculating surprisal are based on the combined original and interpreting corpus for each language, excluding the speech for which surprisal is calculated from the training corpus. This reflects the particular communicative setting in the European Parliament, where a MEP would listen to original speeches and to interpreting production. This modelling approach uses all available spoken corpus data for each language, as the language models for each calculation are already very small in size. However, the subcorpora for interpreting and originals vary slightly in size, particularly in the English dataset, which might lead to a slight bias towards interpreted language. Kunilovskaya et al. (2023b) use the same modelling approach, but a strictly balanced dataset. As their results show

the same trends as the approach used in this thesis, the validity of the approach used here is confirmed.

$$S(w_i) = -log_2 p(w_i|(w_{i-1}w_{i-2}w_{i-3})) \tag{5.3}$$

Surprisal for each word is calculated using the context in which it occurs (cf. Equation 5.2). In the approach used for modelling corpus version EPIC-UdS (aligned, parsed, surprisal), context is defined as the three preceding words (Equation 5.3), which means that surprisal is calculated on a 4-gram basis. Furthermore, (artificially inserted) sentence boundaries are respected, and hapax legomena are replaced with a placeholder string. Words are defined as tokenised strings, which means that different word forms of the same lemma are treated as individual words.

In this thesis, surprisal is used to approximate processing costs for the listener, i.e. processing costs for the interpreter when listening to the original speech as well as processing costs for the recipient of the interpreted speech. Lower surprisal is considered as lower processing costs, and therefore speeches with lower surprisal are defined as more simplified. This is approximated by comparing average surprisal (AvS) (Equation 5.4) for the different subcorpora.

$$AvS(w) = \frac{1}{|w|}\sum_{i=1}^{n} -log_2 p(w_i|(w_{i-1}w_{i-2}w_{i-3})) \tag{5.4}$$

The exploratory analysis of distinctive words in Chapter 4 shows a difference in the use of verbs and nouns when comparing original speeches and interpreting. Verbs were seen as characteristic for interpreting, nominal POS for originals. Therefore, the information content for these categories is further investigated. Surprisal for each word with POS = AUX and POS = VERB in the corpus (EPIC-UdS (aligned, parsed, surprisal)) is extracted. AvS is then calculated for each production mode (original or interpreting) as shown in Equation 5.4. Furthermore, AvS is calculated for POS = NOUN within the different subcorpora. Although the general focus of this thesis is on the language pair German and English (German source speeches interpreted into English and English source speeches interpreted into German - this approach is used in Sections 4.2 and 5.1), this section also considers English interpreting data from Spanish source speeches in the English comparable dataset. This allows for verifying whether effects found in interpreted English are a result of German source structure shining-through or rather a general interpreting effect. Furthermore, the comparable analysis includes effects of delivery mode and looks into individual findings of the exploratory analysis via relative entropy (Section 4.2).

A comparable approach is taken to answer **RQ 2a**, asking whether simultaneous interpreting is more simplified at the lexical level than comparable originals in the same language, including possible influencing factors such as language pair, delivery mode, and distinctive words for originals or interpreting.

In a qualitative analysis, a parallel approach is used to explore triggers of low and high surprisal lexical items in interpreters' output, as well as to investigate how interpreters render low and high surprisal lexical items of the source speeches in their output (**RQ 2b**). Instead of looking at average surprisal, this approach looks into both ends of the surprisal range and particularly selects low and high surprisal verbs and nouns, as distinctive POS resulting from the exploratory analysis (cf. Section 4.3). In EPIC UdS (aligned, parsed, surprisal), surprisal values range from 0.02 to 16.48 in English and from 0.02 to 15.88 in German. For a parallel analysis at the low end of the range, verbs and nouns with surprisal values from 2.0 to 2.9 are filtered, and the most frequent lexical verbs and nouns are selected for further analysis. For the high end of the surprisal range and particularly unexpected words in context, lexical verbs and nouns with surprisal higher than 16.0 are extracted for English and higher than 15.0 for German. In a next step, 50 randomly selected high surprisal verbs and nouns are manually aligned with source and target expressions and investigated further.

5.2.2 Surprisal analysis: general

Although the focus of the analysis of lexical simplification via surprisal is on the distinctive POS resulting from Section 4.2, AvS for all tokens in the individual subcorpora is analysed in a first step to better understand the overall effects on simplification. Table 5.9 and Figure 5.3 show that all the interpreting subcorpora (SI DE EN, SI ES EN and SI EN DE) use words that are more expected in the corresponding contexts than comparable originals (ORG EN and ORG DE). This is indicated by lower AvS (*mean* in the table, visualised as a diamond in the boxplots), but also by lower median surprisal for interpreting compared to original speech (*median* in table, visualised as centre line in boxplots). The trend is more pronounced for the English subcorpora, but also significant for the German subset[13].

subcorpus	N	AvS (mean)	SD	median	IQR	range
ORG EN	64921	6.726913	4.243731	5.92	6.23	16.48
SI DE EN	54323	6.308641	4.179821	5.55	6.07	16.48
SI ES EN	50410	6.226124	4.216177	5.39	6.17	16.48
ORG DE	55550	7.240993	4.175521	6.44	6.32	15.88
SI EN DE	53816	7.150104	4.049804	6.40	5.89	15.88

Table 5.9: Surprisal for all tokens.

To ensure that the picture is not distorted by the written influence of prepared and read-out speeches, delivery rate of source speeches is considered and interpreting

[13]Welch Two Sample t-test for parametric data: ORG EN vs. SI DE EN: t = 17.09, df = 116144, p-value < 2.2e-16; ORG EN vs. SI ES EN: t = 19.951, df = 108688, p-value < 2.2e-16; ORG DE vs. SI EN DE: t = 3.6543, df = 109364, p-value = 0.000258

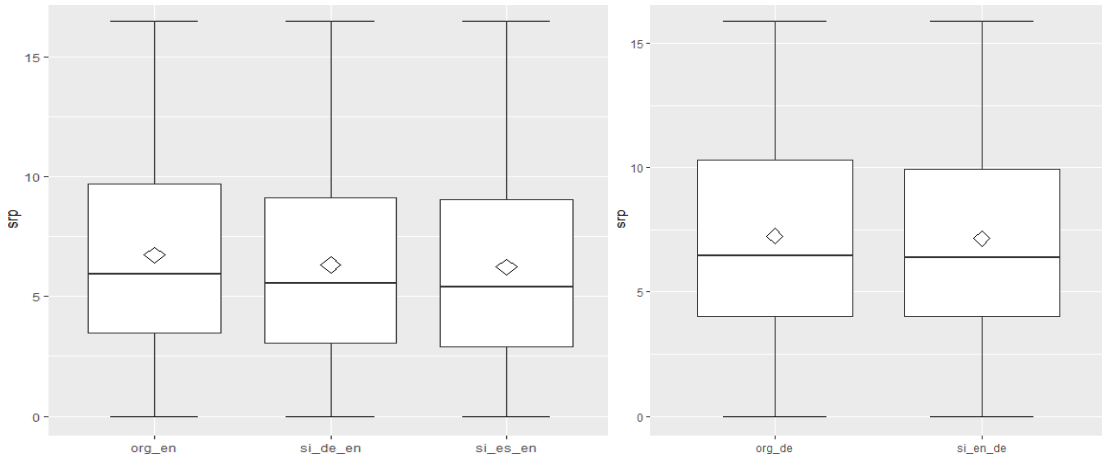


Figure 5.3: Surprisal for all tokens in English (left) and German (right) comparable data.

is compared only to impromptu original speeches. Table 5.10 takes into account the delivery mode of the original speeches, i.e. read, mixed, or impromptu source speech delivery mode. Interpreting is also considered as impromptu delivery type. As expected, prepared and read-out speeches show more variation in the preceding context, indicated by less expected words (higher surprisal) on average. More importantly, impromptu originals also use higher surprisal words than simultaneous interpreting for all comparable data (ORG EN vs. SI DE EN, ORG EN vs. SI ES EN and ORG DE vs. SI DE EN). This difference is statistically significant, although for the German subset, the p-value of 0.04396 is very close to the significance threshold of 0.05[14].

subcorpus	N	AvS (mean)	SD	median	IQR	range
ORG EN read	11642	7.143030	4.425989	6.29	7.12	16.48
ORG EN mixed	38475	6.614432	4.205008	5.79	6.13	16.48
ORG EN impromptu	14804	6.692010	4.177215	5.89	5.97	16.48
SI DE EN (impromptu)	54323	6.308641	4.179821	5.55	6.07	16.48
SI ES EN (impromptu)	50410	6.226124	4.216177	5.39	6.17	16.48
ORG DE read	19227	7.359173	4.293223	6.53	6.76	15.88
ORG DE mixed	18473	7.137229	4.131366	6.39	6.16	15.88
ORG DE impromptu	17850	7.221081	4.088551	6.47	6.00	15.88
SI EN DE (impromptu)	53816	7.150104	4.049804	6.40	5.89	15.88

Table 5.10: Surprisal for all tokens by delivery type.

This result means, by producing more expected utterances overall, interpreters – regardless of source language – produce target speech that is lexically simpler than

[14]Welch Two Sample t-test for parametric data: ORG EN impromptu vs. SI DE EN: t = 9.8976, df = 23506, p-value < 2.2e-16; ORG EN impromptu vs. SI ES EN: t = 11.906, df = 24345, p-value < 2.2e-16; ORG DE impromptu vs. SI EN DE:t = 2.0146, df = 30292, p-value = 0.04396

comparable originals speeches. From a pure lexical perspective, listeners of the interpreted speech would require lower processing processing effort than the audience of original speeches. As observed in Section 4.2 and when analysing lexical density (Section 5.1.1), original and interpreting subcorpora differ in the distribution of function words vs. lexical words. A higher proportion of function words in interpreting may be one possible explanation for the difference in surprisal for this comparison, as function words are generally expected to have lower surprisal. However, there may also be a difference in the lexical items used in general or differences in the context in which a lexical item is used. Therefore, the following sections will analyse individual POS and their context in more detail.

5.2.3 Surprisal analysis: verbs

The method of relative entropy found verbs more typical for interpreting than for original speech. This was found mainly for auxiliary and modal verbs (combined in the universal dependency POS tag AUX), but also lexical verbs (POS tag VERB) were found characteristic for interpreting. Both, AUX and VERB are interesting to investigate further with regard to surprisal.

AUX

Auxiliaries and modal verbs belong to the closed word classes, and therefore interpreters and original speakers will use the same (limited) list of linguistic items. However, the context and, therefore, the (un)expectedness may differ. If the simplification hypothesis holds, the contexts in which auxiliary and modal verbs are used in interpreting will be more predictive, resulting in lower values of surprisal for AUX in interpreting, and higher surprisal in originals.

subcorpus	N	AvS (mean)	SD	median	IQR	range
ORG EN	4689	4.931764	2.617985	4.63	3.830	14.62
SI DE EN	4623	4.672637	2.597365	4.24	3.885	13.96
SI ES EN	3656	4.481335	2.521157	4.02	3.690	14.62
ORG DE	3830	6.279287	2.916184	6.29	4.260	15.76
SI EN DE	4358	6.127324	3.007841	6.28	4.668	15.73

Table 5.11: Surprisal for auxiliary and modal verbs.

Table 5.11 and Figure 5.4 show that interpreters use AUX in more constrained context than original speakers, indicated by lower AvS. For both languages, significant differences between originals and interpreting are confirmed by a Welch Two Sample t-test for parametric data[15]. This observation also holds for the English dataset with

[15]ORG EN vs. SI DE EN: t = 4.7946, df = 9309.6, p-value = 1.656e-06; ORG EN vs. SI ES EN:

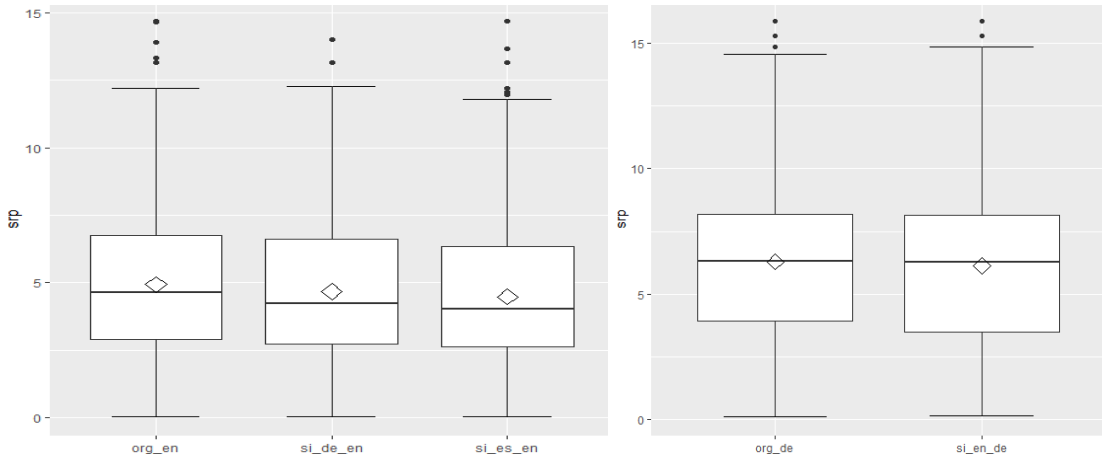


Figure 5.4: Surprisal for auxiliary and modal verbs in English (left) and German (right) comparable data.

AvS lower in interpreted English from both source languages (German and Spanish) than in the comparable English original subcorpus.

To make sure that this is not just a difference in preparedness – due to written influence in the prepared original speeches – delivery mode is considered as an influencing variable and simultaneous interpreting is compared directly to impromptu original speeches. The results are shown in Table 5.12. The expected trend is confirmed for English when interpreting from two source languages, where auxiliary and modal verbs in simultaneous interpreting (considered as impromptu delivery type) are significantly more expected than in impromptu original speeches. For German, the same trend can be observed, but the difference is not significant[16].

As expected, the difference between read-out and prepared speeches compared to simultaneous interpreting is even clearer. Therefore, preparedness does influence surprisal: If a speech is prepared, it includes more variation of n-grams used (surprisal is here calculated using a context of 3 tokens). Spontaneous production, and interpreting in particular, uses auxiliary and modal verbs in more standardised, and therefore expected contexts. This makes processing of lexical input for the audience of the interpreted speech easier and the interpreting product is simpler as a result. However, this is not seen as an audience design mechanism by the interpreter. The declining surprisal rate from read to impromptu original speeches, to even lower surprisal in interpreting shows that this is rather a production constraint, from written-influenced production to spoken, with simultaneous interpreting being more spoken than spoken originals.

t = 7.9622, df = 7984.3, p-value = 1.923e-15; ORG DE vs. SI EN DE: t = 2.3184, df = 8107.6, p-value = 0.02045

[16]Welch Two Sample t-test for parametric data: ORG EN impromptu vs. SI DE EN (impromptu): t = 3.2887, df = 1832.9, p-value = 0.001026; ORG EN impromptu vs. SI ES EN (impromptu): t = 5.4074, df = 1949.6, p-value = 7.18e-08; *ORG DE impromptu vs. SI EN DE (impromptu): t = 0.85152, df = 2274.4, p-value = 0.3946 - not significant

subcorpus	N	AvS (mean)	SD	median	IQR	range
ORG EN read	706	5.173952	2.557529	5.00	3.7975	14.36
ORG EN mixed	2780	4.859723	2.603019	4.55	3.9000	13.78
ORG EN impromptu	1203	4.956110	2.679997	4.62	3.8050	14.62
SI DE EN (impromptu)	4623	4.672637	2.597365	4.24	3.885	13.96
SI ES EN (impromptu)	3656	4.481335	2.521157	4.02	3.690	14.62
ORG DE read	1120	6.441134	2.891729	6.67	4.3625	15.04
ORG DE mixed	1354	6.218257	2.860795	6.29	4.2275	15.73
*ORG DE impromptu	1356	*6.206549	2.987109	6.29	4.7425	14.76
*SI EN DE (impromptu)	4358	*6.127324	3.007841	6.28	4.668	15.73

Table 5.12: Surprisal for auxiliary verbs by delivery type.

A more detailed investigation of individual auxiliaries shows that highly surprising auxiliaries occur more frequently in original speech than interpreting: The English auxiliary *may* (AvS = 9.15) occurs with a frequency per million words (fpm) of 0.0724 in originals vs. fpm of 0.0239 in SI DE EN and fpm of 0.0317 in SI ES EN. *shall* (AvS = 12.4) is not used in SI DE EN at all, but in originals with fpm of 0.0061 and in SI ES EN with fpm of 0.0060. However, these highly surprising auxiliaries are not used very frequently overall and, therefore, cannot be the main driver of average surprisal being lower for AUX in SI vs. ORG overall. The high-frequent auxiliaries make a difference on a large scale. *be* as the most frequent English auxiliary in all subsets[17] is used in more expected context in interpreting (SI DE EN AvS = 3.73, SI ES EN AvS = 3.59) than in originals (AvS = 4.05)[18]. Example (23) shows *be* used in an if-clause in original English, a high surprisal use of this auxiliary, whereas in the example taken from the interpreting subcorpus (24), *be* is used in expected context (*...and this debate is...* as well as in contexts like *...that this decision is*, which occur several times in the corpus).

(23) ...if the deadlines were (9.68) not met...

(24) ...and this decision is (2.48) going to help...

The most frequent auxiliary in German is *sein*, occurring more frequently in interpreting than in originals[19]. *sein* occurs in more expected contexts in simultaneous interpreting than in originals (ORG DE AvS = 5.50; SI EN DE AvS = 5.42), how-

[17]fpm in ORG EN: 4.1312; in SI DE EN: 5.0181 and in SI ES EN 4.2412

[18]All differences in AvS ORG vs. SI are statistically significant, confirmed by a Welch Two Sample t-test for parametric data.

[19]fpm in ORG DE: 25,400.54 and in SI EN DE 25,884.50

ever, this trend is – in line with the general trend for German in this category – not significant[20].

To summarise the results for auxiliary and modal verbs, it can be said that interpreting generally uses lower average surprisal AUX than original speeches. This can be observed for both comparable datasets (English and German) and is more pronounced in English. Results of the English comparable dataset show that this trend is observable for English interpreting from two different source languages, pointing to an interpreting effect, rather than a source language influence. Furthermore, delivery mode has an influence on lexical variation with AvS with lower in interpreting than impromptu originals. This points to interpreting to be more spoken than original spoken. For AUX as closed word class, a frequency effect can also be observed. Although interpreters and original speakers use the same limited list of linguistic items, interpreters use the same auxiliaries and modals more frequently and in more standardised contexts.

VERB

Lexical verbs (universal dependency POS tag VERB), on the other hand, are an open word class. In the exploratory analysis (cf. Section 4.2), it was observed that interpreting is characterised by the use of mostly very general lexical verbs and therefore the vocabulary interpreters and original speakers use may differ. Querying POS VERB and extracting surprisal of these verbs in the different subcorpora gives the following result.

As in the analyses for all tokens and AUX above, average surprisal for this word class also shows a clear trend of interpreters using lexical verbs that are more expected in context than comparable original speakers of both languages studied. For English, this trend can also be observed for interpreting from both source languages (cf. Table 5.13 and Figure 5.5)[21]. The trend, again, is clearer for English comparable data, but also significant for the German comparison.

Considering only impromptu speeches in the analysis gives a slightly different picture. For English, lexical verbs in interpreting from two different sources languages (German, Spanish) are again more expected (lower AvS) than impromptu original speeches. The same tendency can be observed for German, too; however, the difference is, like with the above-studied auxiliaries and modal verbs, not statistically significant[22].

[20]Wilcoxon rank sum test with continuity correction: W = 1000755, p-value = 0.4011 - not significant

[21]Welch Two Sample t-test for parametric data: ORG EN vs. SI DE EN t = 13.003, df = 14493, p-value < 2.2e-16; ORG EN vs. SI ES EN t = 13.175, df = 13767, p-value < 2.2e-16; ORG DE vs. SI EN DE t = 2.0176, df = 6827.8, p-value = 0.04367

[22]Welch Two Sample t-test for parametric data: ORG EN impromptu vs. SI DE EN (impromptu): t = 7.4304, df = 3022.5, p-value = 1.402e-13; ORG EN impromptu vs. SI ES EN (impromptu): t

subcorpus	N	AvS (mean)	SD	median	IQR	range
ORG EN	8066	8.851603	4.403067	8.74	7.37	16.45
SI DE EN	6841	7.908724	4.419294	7.45	7.28	16.46
SI ES EN	6462	7.876154	4.459730	7.45	7.39	16.46
ORG DE	2495	10.018377	4.452897	11.32	8.135	15.86
SI EN DE	3145	9.744898	4.376439	10.59	8.120	15.86

Table 5.13: Surprisal for lexical verbs.

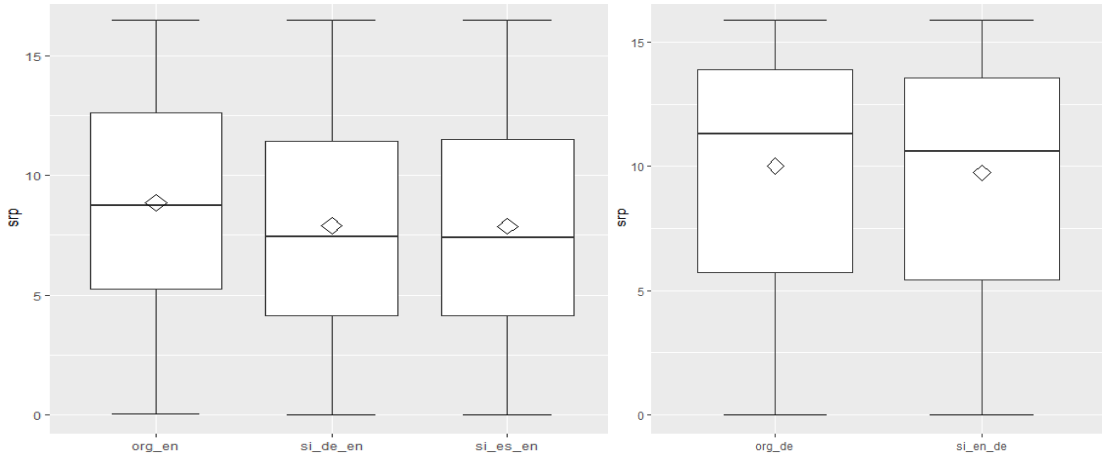


Figure 5.5: Surprisal for lexical verbs in English (left) and German (right) comparable data.

subcorpus	N	AvS (mean)	SD	median	IQR	range
ORG EN read	1375	9.731607	4.515662	10.180	7.82	16.45
ORG EN mixed	4813	8.638186	4.359551	8.430	7.21	16.43
ORG EN impromptu	1878	8.754249	4.353913	8.775	7.18	16.43
SI DE EN (impromptu)	6841	7.908724	4.419294	7.45	7.28	16.46
SI ES EN (impromptu)	6462	7.876154	4.459730	7.45	7.39	16.46
ORG DE read	1681	10.120048	4.483218	11.39	8.26	15.86
ORG DE mixed	1469	9.766188	4.382914	10.93	8.17	15.86
*ORG DE impromptu	1296	*9.944066	4.440186	10.99	8.37	15.84
*SI EN DE (impromptu)	3145	*9.744898	4.376439	10.59	8.12	15.86

Table 5.14: Surprisal for lexical verbs by delivery mode.

Analysing results of relative entropy The method of relative entropy showed individual lexical verbs distinctive for interpreting. These include *talk* and *react* for English interpreting and *sicherstellen, arbeiten, sagen, geben* and *freuen* for German

= 7.651, df = 3112.1, p-value = 2.642e-14; *ORG DE impromptu vs. SI EN DE: t = 1.3648, df = 2382.2, p-value = 0.1724 - not significant

interpreting. As the dataset is generally very small, surprisal is extracted for these word lemmas, and AvS is compared. The results follow the general observed trend with AvS lower in interpreting than in interpreting[23].

Looking at distinctive lexical verbs for English interpreting, the following observation can be made. AvS for all lemmas of *talk* is significantly higher in original English speeches (AvS = 8.336) compared to *talk* in interpreting from German into English (AvS = 6.367661) as well as from Spanish into English (AvS = 6.16619)[24]. The second distinctive lexical verb for interpreting *react* is very unexpected (high AvS) in interpreting from both source languages (SI DE EN AvS = 12.857; SI ES EN AvS = 10.450) and does not occur in original English speeches.

The same trend can also be observed for lexical verbs that the method of relative entropy found typical for English original speeches. *believe* is more expected in interpreting (SI DE EN AvS = 6.865000; SI ES EN AvS = 7.743929) than original speeches (AvS = 8.051186). *recognise* on the other hand is not used in interpreting at all, but occurs in unexpected context in the original dataset (AvS = 12.25923).

When preparedness of original speeches is considered, the same general trend can be observed. Figure 5.6 shows the same trend for lexical verb *talk* when only impromptu original speeches are considered: AvS for *talk* in ORG EN impromptu (AvS = 7.662) compared to AvS in interpreting from both source languages is lower (SI DE EN AvS = 6.368; Figure 5.6 on the left and SI ES EN AvS = 6.116; Figure 5.6 on the right).

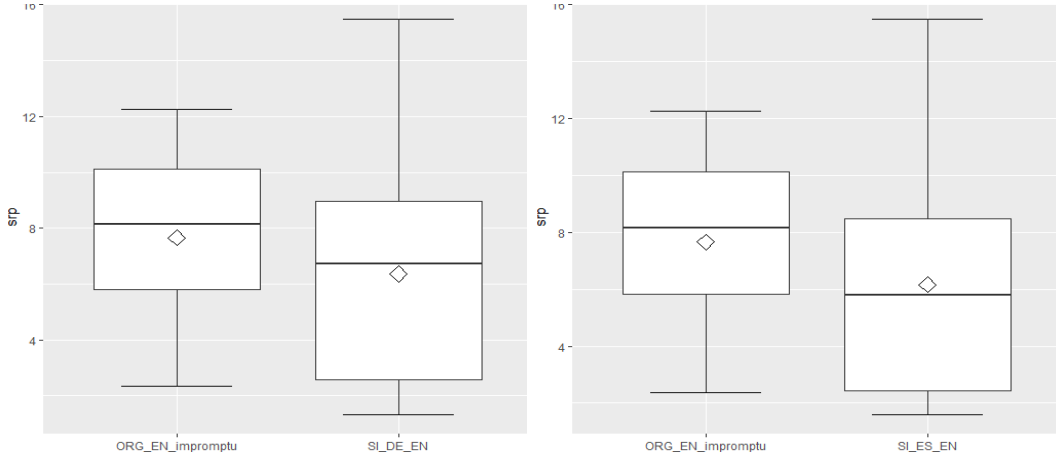


Figure 5.6: Surprisal for *talk* for impromptu originals vs. interpreting from German (left) and from Spanish (right).

A more detailed investigation of low surprisal (surprisal of 2.++) and high surprisal (surprisal of 8 and higher) occurrences of *talk* shows that low surprisal *talk* occurs

[23]Basic statistics for the individual verbs are given in the Appendix (Table A.5)

[24]Welch Two Sample t-test for parametric data: ORG EN vs. SI DE EN: t = 3.6445, df = 103.72, p-value = 0.0004205; ORG EN vs. SI ES EN: t = 3.7468, df = 113.49, p-value = 0.0002832

mainly in both interpreting subcorpora (SI DE EN with fpm = 515.44 and SI ES EN with fpm = 456.26), whereas it is rare in such low surprisal context in ORG EN (fpm = 92.42). All of these low surprisal occurrences in interpreting are the use of *talk* in relative clauses as present progressive in the first person plural (see example 25). In interpreting, high surprising contexts of *talk* is mostly preceded by *we* or the second person singular pronoun *you*, followed by *talk* in different tenses. In original speeches, the agents and the syntax used are more diverse in the original speeches (example 26). Therefore, even in very unexpected contexts, more variation can be observed for originals than for interpreting.

(25) ...that's exactly the type of thing that we're *talking* (srp = 2.58) about...

(26) ...as we know with the midwives recently *talking* (srp = 10.33) to about...

Although AvS of lexical verbs for impromptu modes does not differ significantly in the German subsets, it is still interesting to investigate the distinctive lexical verbs resulting from the analysis in Section 4.2 and see whether the general trend can also be observed for individual verbs. *sicherstellen, arbeiten, sagen, geben* and *freuen* were found to be distinctive for interpreting by means of relative entropy, while only *reden* and *setzen* are characteristic for German originals.

The most distinctive lexical verb for interpreting *sicherstellen* is used in highly standardised context in this mode (cf. example 27 and 28), following in first person plural and a modal verb (AvS = 8.27). On the other hand, *sicherstellen* only occurs 4 times in original German speeches and only once in impromptu delivery mode (cf. example 29). The occurrence in the impromptu speech as well as another occurrence use *sicherstellen* in a subordinate clause (infinitive clause in the impromptu example). The other occurrence for original speech use of *sicherstellen* is as passive (example 30). This use is unexpected in context, resulting in average surprisal for *sicherstellen* in originals of 14.37, which is much higher than in interpreting (AvS = 8.27).

(27) ...wir müssen *sicherstellen* (srp = 4.45) dass ...

(28) ...zweitens müssen wir *sicherstellen* (srp = 6.22) dass ...

(29) ...wir müssen auch rangehen die Verlängerung *sicherzustellen* (srp = 12.84) ...

(30) ...Nachhaltigkeit der Sozial- und Pensionssysteme sollten *sichergestellt* werden (srp = 14.88) ...

A similar picture can be observed for the other lexical verbs that were found to be distinctive for interpreting: *arbeiten* is more expected in interpreting (AvS = 10.32) than in impromptu originals (AvS = 12.65). The same trend can be observed for *geben* (AvS SI = 7.64 vs. AvS ORG impromptu = 7.82) and *sagen* (AvS SI = 7.20 vs. AvS ORG impromptu = 7.93), although statistical significance can only be proven for *sagen*[25]. For the lexical verbs that were distinctive for originals, *reden* also occurs in more expected context in interpreting (AvS SI = 10.60 vs. AvS ORG impromptu = 10.74), although not significantly. The trend is slightly reversed for *setzen* with AvS = 13.66 for interpreting and AvS = 13.50 for impromptu originals; however, the difference is not statistically significant.

With respect to analysing surprisal in the results of the exploratory analysis (relative entropy), it can be concluded that individual lexical verbs, no matter whether they are distinctive for interpreting or originals, show a tendency to have lower average surprisal in interpreting than originals and therefore occur in more expected context in interpreting. However, this is again more pronounced in English and is only statistically significant for some of the lexical verbs analysed in German.

Analysing potential triggers as well as translations of low surprisal lexical verbs The previous analyses of surprisal for lexical verbs in the sections above showed a trend of lexical verbs in general as well as selected lexical verbs to be more expected in interpreting vs. comparable original speech of the same language. The following analysis tries to investigate possible simplification trends from the source to the target and to look into potential triggers of such expected use of verbs via a parallel approach: low surprisal lexical verbs in interpreting and their triggers in their source (original speech) are investigated. Furthermore, to verify whether these results are the result of the interpreting process, or whether the same trend is observed for the language pair in general, low surprisal lexical verbs in original speeches and their translations in the target (interpreting) are also investigated. In a first step, all low surprisal lexical verbs with surprisal values of 2 (2.0 to 2.9) in the parallel corpora ORG EN aligned with SI EN DE and ORG DE aligned with SI DE EN are extracted (cf. Table A.7 in the Appendix). The resulting list also includes several verbs that are ambiguous in their use, such as *be*, which can be used as an auxiliary or lexical verb. Due to their ambiguity in use and the potential for tagging errors in this assignment, these verbs are not considered further in this analysis. Interestingly, the second most frequent low surprisal lexical verb in both English subcorpora is the lexical verb *think*. *glauben* as a possible translation thereof is also at the top of the list of most frequent low surprisal lexical verbs in German (second most frequent in ORG DE and third most frequent in SI EN DE). Therefore, low surprisal occurrences of these two verbs as well as their corresponding source or target text items and their surprisal are

[25]Welch Two Sample t-test: t = 2.0488, df = 169.59, p-value = 0.04202

extracted (Table 5.16)[26]. Surprisal values across two languages can generally not be compared directly. However, since the starting point of this analysis is a range of low surprisal values in both languages (2.0 to 2.9) and the relative frequencies across both languages are comparable, average surprisal (AvS) of the corresponding source or target items are linked. At the same time, the analysis takes into account that AvS for lexical verbs in general was found to be lower in English than in German (cf. Table 5.13 and Figure 5.5 above).

In general, it can be observed that low surprisal verbs are much more frequent in interpreting than the comparable original dataset, with a clearer difference for English than for German (cf. Table 5.15 for raw frequencies (rf) and relative frequencies (fpm) of low surprisal lexical verbs). This confirms the observed tendency above of interpreting using expected verbs more frequently.

ORG EN		SI DE EN		ORG DE		SI EN DE	
rf	fpm	rf	fpm	rf	fpm	rf	fpm
483	7,439.81	598	11,008.23	169	3,042.30	183	3,400.48

Table 5.15: Frequencies for low surprisal lexical verbs with spr = 2.0 - 2.9.

The low surprisal parallel analysis of the selected lexical verbs gives the following results. Low surprisal (I)[27] *think* in the English original speeches is often left out in interpreting into German (65.1% of cases). These cases are mainly the use of (I) think as a pragmatic discourse marker. If the low surprisal source text *think* is translated into German, it is translated as *glaube* (18.1%) or *denke* (10.8%), also with fairly low AvS for these verbs in German (considering that lexical verbs in German have slightly higher AvS overall). Other translations of low surprisal *think*, such as translations with stance adverbials *sicher* (srp = 11.32) or *wahrscheinlich* (spr = 13.3) are rare with only one occurrence. Low surprisal *glauben*, on the other hand, is mostly translated by a lower surprisal target rendition of *think* (66.7%) or is also left out in the target production (27.8%). It can be summarised for these low surprisal items in the source that they are mostly either easy to be processed by the listener (rendered with a low surprisal target expression) or left out completely and therefore do not have to be processed at all by the listener of the interpretation.

Looking at low surprisal *think* and *glaube* in the target and their triggers in the source, a similar trend can be observed: (ich) *glaube* is mostly triggered by an equally low surprising *think* (66.7%). However, in 25% of cases, low surprisal *glaube* is the translation for a higher surprisal source expression such as *suspect* with srp = 8.35 or

[26]The distinction of speeches with different delivery type is not made for the parallel investigation as the phenomena investigated are low frequency and therefore sparsity of data becomes an issue when considering only impromptu original speeches.

[27]Surprisal values for *ich* and *I* are not included in the calculation, but are mentioned in brackets as these pronouns are always in the preceding context for the low surprisal use of these verbs.

ORG EN (source *think*, spr = 2.++)	SI EN DE (target expressions)
83 x (I) think	15 x (18.1%) (ich) glaube, AvS = 4.73 9 x (10.8%) (ich) denke, AvS = 5.79 54 x (65.1%) none 5 x (6.0%) misc, AvS = 8.22
ORG EN (source triggers)	SI EN DE (target *think*, spr = 2.++)
8 x (66.7%) think, AvS = 2.89 1 x (8.3%) none 3 x (25.0%) misc, AvS = 7.89	12 x (ich) glaube
ORG DE (source *glauben*, spr = 2.++)	SI DE EN (target expressions)
18 x (ich) glaube	12 x (66.7%) think, AvS = 1.88 5 x (27.8%) none 1 x (5.6%) misc (believe srp = 5.04)
ORG DE (source triggers)	SI DE EN (target *glauben*, spr = 2.++)
24 x (24.7%) glauben, AvS = 5.27 7 x (7.2%) denken, AvS = 5.05 45 x (46.4%) none 21 x (21.7%) misc, AvS = 7.34	97 x (I) think

Table 5.16: Low surprisal *think/glauben* and corresponding source/target expressions.

doubt with srp = 11.20. (I) *think*, on the other hand, is mostly used as low surprisal filler or pragmatic discourse marker in English interpreting. In 46.4% of cases, there is no trigger in the source speech. If the trigger in the source speech is *glauben* or *denken*, it is still fairly expected (considering higher average surprisal for lexical verbs in German overall). Among the divers triggers are several equally low surprisal triggers (e.g. *hoffe* srp = 4.95), but mostly triggers that are unexpected and therefore more difficult to process for the interpreter, such as stance adverbial *eigentlich* (srp = 10.39) or more complex expressions such as *habe ich den Eindruck* (spr = 10.18 + 2.58 + 5.33 + 0.89). Such multi-word expressions are not necessarily harder to process on average (AvS = 4.75), however, multiple items have to be processed which increases the processing load overall.

Low surprisal *(I) think* is mostly used as a pragmatic discourse marker in original and interpreted English. Previous research also showed that interpreters use discourse markers independently of the source and tend to leave out discourse markers from the source or add them to their target production (Defrancq, 2016). This tendency can be observed in both directions in this analysis. By leaving out low surprisal discourse markers, the listener does not need to process the filler and therefore is processing fewer individual items. This may lead to higher information density of the message in general, increasing total processing load. However, AvS of all tokens in interpreting was lower than in original speech (cf. Table 5.9); therefore, it can

be assumed that leaving out low surprisal discourse markers in interpreting does not increase processing load overall. This trend cannot be observed for German where *ich glaube* is not frequently added without a trigger in the source.

To summarise the investigation of low surprisal lexical verbs and their corresponding source and target expressions, it can be concluded that low surprisal verbs can be linked mostly to corresponding low surprisal verbs, or to no corresponding (source or target) item at all. This means that processing effort for these low surprisal items is also low or no processing effort is required at all (in case of left-out low surprisal items). This can be observed for low surprisal source text lexical verbs and their interpretation as well as for low surprisal lexical verbs in interpreting and their triggers in the source. The main difference between the two comparisons (translations and triggers of low surprisal verbs) can be found in the *misc* renditions (Table 5.16). While it is very rare that a low surprisal lexical verb in originals is rendered with a high surprisal lexical verb, phrasal verb, as nominal phrase or other (6.0% and 5.6% for German and English target expressions), such unexpected and longer source text triggers are often simplified in interpreting by using a low surprisal lexical verb (25.0% and 21.7% of cases for English and German source triggers, respectively). The simplification hypothesis can therefore be confirmed for this part of the investigation.

Analysing potential triggers and translations of high surprisal lexical verbs To further investigate simplification from the source to the target and explore how very unexpected lexical verbs of the source speech are rendered in the target speech as well as how high surprisal lexical verbs of the target are triggered, lexical verbs with highest surprisal values in the respective datasets are investigated further. Surprisal for lexical verbs ranges from 0.02 to 16.48 in English and from 0.02 to 15.88 in German. To select highest surprisal items in each language, lexical verbs with surprisal higher than 16.0 are extracted for English and higher than 15.0 for German. Table 5.17 shows raw frequencies (rf) and relative frequencies (fpm) of these high surprisal verbs for the different subcorpora.

ORG EN		SI DE EN		ORG DE		SI EN DE	
rf	fpm	rf	fpm	rf	fpm	rf	fpm
290	4,466.97	354	3,380.02	648	11,665.17	608	11,297.76

Table 5.17: Frequencies for high surprisal lexical verbs with spr $>$ 16 in English and srp $>$ 15 in German.

A first general observation is that there are major frequency differences for both languages. High surprisal lexical verbs occur much more frequently in German than in English, even though lexical verbs are more frequent in general. Overall, AvS for lexical verbs was also found to be higher in German than English (see this section above). As surprisal for this corpus is calculated on the basis of strings or word forms,

not lemmas, German's richer morphology and more inflected forms for each lexical word may result in higher surprisal for individual verbs. However, a difference between originals and interpreting can also be found in both languages. These high surprisal lexical verbs occur more frequently in originals than in interpreting, pointing again to less variation preceding context in interpreting and therefore simpler interpreter's output.

Source and target speeches in the corpus are aligned at the sentence level. However, corresponding source or target items of these high surprisal lexical verbs have to be aligned manually, and corresponding surprisal values of such items have to be extracted individually. In order to get a representative but manageable sample, a subset of 50 randomly selected high surprisal lexical verbs for each source and target language are investigated in more detail. The following categories for aligning source and target item were chosen. *Lexical verb* for aligned items of high surprisal lexical verbs with formal equivalence: the corresponding translation is within the same grammatical category, i.e. also a lexical verb. The focus for alignment is on formal equivalence, not on semantic equivalence, e.g. DE *betonen* (stress or highlight) - EN: *echo* is seen as formally equivalent although there may be semantic differences. The category *lexical verb* is the only category that is compared directly and for which AvS is calculated[28].

Hapax legomena lexical verbs are treated as a separate category as their assigned surprisal values in the corpus do not reflect actual processing difficulty[29]. If no corresponding source or target item was found – either the high surprisal lexical verb was left out or the entire sentence was not aligned – the category chosen is *none*. The category *other* is used for any kind of transposition: lexical verb being aligned to a noun phrase (e.g. (was ...) *anbelangt* - *the aspect of*; *drafted* - *Verfasserin*), adjective (e.g. *verified* - *kontrollierbar*) or other. Table 5.18 gives an overview of how the high surprisal verbs of the source were rendered in the interpreter's output and how high surprisal verbs in interpreting were triggered in the source.

The results for high surprisal lexical verbs and their corresponding triggers or translations are not as clear as for the low surprisal analysis. High surprisal lexical verbs in the source are mostly also translated with a lexical verb in the target (48% of cases for ORG EN to SI EN DE and 40% in ORG DE to SI DE EN). The lexical verbs in the source were extracted from the corpus as verbs from the top end of the range for surprisal within the respective subcorpora, whereas the lexical verbs to which this source input translates are below the top end of the range for surprisal. For the direction ORG EN into SI EN DE, average surprisal for the target lexical verbs is

[28]In case the aligned lexical verb is a phrasal verb: surprisal was taken only from the verb, not the preposition, e.g. *voted* (against) - (dagegen) *gestimmt*.

[29]Unique strings theoretically should have the highest possible surprisal within the corpus. As they only occur once, they are most unexpected in any context. However, in the corpus variant used, unique strings are combined in one bucket, which is then treated as one string with many occurrences. Therefore, these unique strings have lower surprisal values than their real surprisal value.

ORG EN (source srp = 16.++)	SI EN DE (target expressions)
50 x	24 x (48%) lexical verb, AvS = 12.33 6 x (12%) hapax legomena lexical verbs 7 x (14%) none 13 x (26%) other
ORG EN (source triggers)	SI EN DE (target srp = 15.++)
32 x (64%) lexical verb, AvS = 11.06 1 x (2%) hapax legomena lexical verbs 7 x (14%) none 10 x (20%) other	50 x
ORG DE (source srp = 15.++)	SI DE EN (target expressions)
50 x	20 x (40%) lexical verb, AvS = 8.96 2 x (4%) hapax legomena lexical verbs 16 x (32%) none 12 x (24%) other
ORG DE (source triggers)	SI DE EN (target srp = 16.++)
13 x (26%) lexical verb, AvS = 13.61 16 x (32%) hapax legomena lexical verbs 3 x (6%) none 18 x (36%) other	50 x

Table 5.18: High surprisal lexical verbs and corresponding source/target expressions.

much lower than the maximum surprisal (with AvS of 12.33 compared to a maximum value of 15.88 and AvS for this POS in this category of 9.74). For the direction ORG DE into SI DE EN, a clear simplification trend can be observed. Maximum surprisal lexical verbs in the source translate to AvS of 8.96, which is clearly below the maximum of 16.48 and close to AvS of this POS in this category of 7.91 (cf. Tables 5.18 and 5.19).

High surprisal verbs in the source are rarely translated with a uniquely occurring word form (hapax legomena) with only 12 % of cases for ORG EN into SI EN DE and 4% of cases studied for ORG DE into SI EN DE. On the other hand, hapax legomena are often the trigger in the German source speech for high suprisal verbs in SI DE EN (32% of cases).

The main difference can be observed in the category *none* for the translation direction ORG DE to SI DE EN. High surprisal lexical verbs in the German source are often left out in the English interpretation (32% of cases), whereas only 6% of high surprisal verbs in SI DE EN have no trigger in the source. No difference can be observed for ORG EN to DI EN DE with 14% of non-aligned high surprisal verbs for source triggers and target expressions. This points to a simplification trend for the interpreting direction German into English, where items that were difficult to process

in the source speech are not rendered in the target output.

However, the overall relatively high frequency of hapax legomena within this investigation and their incorrect assignment of surprisal values show the limits of the surprisal calculation for this corpus variant again, as the corpus size is relatively small and topics covered diverse[30]. In order to get more reliable results, it would be possible to consider using a bigger reference corpus (with EP speeches of other corpora) in order to get a more comprehensive picture of the language use within the European Parliament. Some of the unique strings within EPIC-UdS are generally not very low frequent lemmas (e.g. *orientierte* or *protesting*). The corpus size of the EPIC-UdS supcorpora for English and German might be not large enough to account for string expectedness in context. Considering lemma surprisal may be one way of dealing with this problem. Furthermore, correcting surprisal for hapax legomena with the highest possible surprisal value for this corpus would give a more correct picture.

ORG EN	SI EN DE (target)	ORG DE (comp. triggers)	all SI EN DE
16.++	AvS = 12.33	AvS = 11.06	AvS = 9.74
ORG DE	SI DE EN (target)	ORG EN (comp. triggers)	all SI DE EN
15.++	AvS = 8.96	AvS = 13.61	AvS = 7.91
ORG EN (triggers)	SI EN DE	SI DE EN (comp. target)	all ORG EN
AvS = 13.61	15.++	AvS = 8.96	AvS = 8.85
ORG DE (triggers)	SI DE EN	SI EN DE (comp. target)	all ORG DE
AvS = 11.06	16.++	AvS = 12.33	AvS = 10.02

Table 5.19: High surprisal lexical verbs, corresponding source/target lexical verbs and average surprisal for comparable data.

To summarise the findings for verbs, it can be stated that a general trend of simplification can be found for both languages, with a greater effect in English than in German. In the comparable approach, AvS for all verbs in interpreting is found to be lower than in comparable originals, however, this trend is only confirmed for English, not for German, when considering delivery mode. From a parallel perspective, a simplification trend can be observed for high and low surprisal verbs. While low surprisal verbs also tend to be translated with low surprisal verbs or are left out completely, a simplification trend can particularly be seen for lexical verbs at the higher end of the surprisal range. Highly unexpected verbs in originals are translated with more expected verbs and are frequently left out in interpreting. At the same time, highly unexpected verbs in interpreting are frequently triggered by hapax legomena verbs. In general, it can be concluded that lexical verbs in interpreting tend to be simpler than the originals on which they were based.

[30]For English 2.13% of all tokens (or 3619 tokens) only occur once, in German, this is true for 6.28% of tokens (or 6865 tokens).

5.2.4 Surprisal analysis: nouns

The method of relative entropy found that nouns are generally more typical for original production than interpreted speech. Only very few general nouns characterise the distributions of interpreted language (cf. Section 4.2). This POS is therefore particularly interesting to investigate as it is more distinctive for originals, whereas the POS looked at above were more distinctive for interpreting. In a first step, surprisal for all nouns in the subcorpora is investigated and average surprisal compared, before analysing noun surprisal in accordance to delivery mode, investigating individual nouns that the method of relative entropy found distinctive for interpreting or originals, and finally taking a parallel approach to investigate simplification in nouns from the source to the target.

Table 5.20 and Figure 5.7 give the results for noun average surprisal in the subcorpora. They show the same trend for original English compared to interpreting into English from both source languages. English interpreting uses significantly more expected nouns than original speech[31], in line with the findings for all tokens, AUX and VERBs investigated above. However, the opposite trend can be observed for the German dataset, with originals showing lower surprisal nouns than interpreting[32]. For German, this is the opposite trend compared to the general analysis of all tokens, AUX and VERBs investigated above.

subcorpus	N	AvS (mean)	SD	median	IQR	range
ORG EN	11878	9.600407	4.387393	10.61	6.99	16.48
SI DE EN	9380	9.171248	4.452860	10.05	7.01	16.48
SI ES EN	9073	9.201752	4.488139	10.07	7.13	16.47
ORG DE	11004	8.573837	4.998693	9.29	9.23	15.88
SI EN DE	9474	8.819415	4.795883	9.45	8.84	11.88

Table 5.20: Surprisal for nouns.

Taking delivery mode into account and only comparing impromptu delivery modes (cf. Table 5.21) confirms the trend that was also found for the other POS investigated above. English interpreting uses significantly more expected nouns than comparable impromptu originals. This is expected as planned and read-out speeches are expected to use more varied and sophisticated vocabulary. However, there are no significant differences for German interpreting vs. comparable impromptu originals[33]. German

[31]Welch Two Sample t-test for parametric data: t = -3.5827, df = 20238, p-value = 0.0003408 ORG EN vs. SI DE EN: t = 7.0227, df = 19993, p-value = 2.246e-12 ; ORG EN vs. SI ES EN: t = 6.4326, df = 19296, p-value = 1.283e-10

[32]Welch Two Sample t-test for parametric data: ORG DE vs. SI EN DE: t = -3.5827, df = 20238, p-value = 0.0003408

[33]Welch Two Sample t-test for parametric data: ORG EN impromptu vs. SI DE EN (impromptu): t = 15.93, df = 4176.7, p-value < 2.2e-16; ORG EN impromptu vs. SI ES EN (impromptu): t =

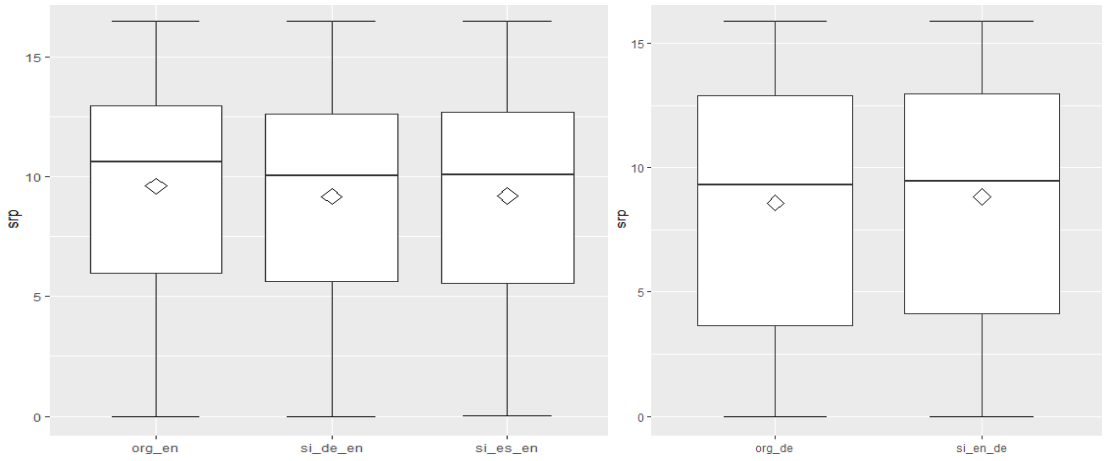


Figure 5.7: Surprisal for nouns in English (left) and German (right) comparable data.

interpreting uses even more unexpected nouns than read originals, the category that in the previous analyses always was highest in surprisal. Therefore, simplification in interpreting can be observed for English nouns, but no difference can be found for German.

subcorpus	N	AvS (mean)	SD	median	IQR	range
ORG EN read	2538	10.005260	4.326393	11.07	6.77	16.47
ORG EN mixed	6939	9.457246	4.380502	10.33	7.06	16.47
ORG EN impromptu	2401	9.586193	4.446237	10.56	7.17	16.48
SI DE EN (impromptu)	9380	9.171248	4.452860	10.05	7.01	16.48
SI ES EN (impromptu)	9073	9.201752	4.488139	10.07	7.13	16.47
ORG DE read	4224	8.562488	5.097840	9.35	9.53	15.88
ORG DE mixed	3532	8.465048	4.998391	9.02	9.21	15.88
*ORG DE impromptu	3248	*8.706897	4.865484	9.36	8.72	15.88
*SI EN DE (impromptu)	9474	*8.819415	4.795883	9.45	8.84	11.88

Table 5.21: Surprisal for nouns by delivery mode.

Analysing results of relative entropy As nominal POS are more distinctive for originals, it is interesting to further investigate nouns that the method of relative entropy found to be distinctive for of each subset and find out if nouns behave differently than other POS that were found to be distinctive for interpreting. The distinctive nouns (cf. Section 4.2) include many topic-related nouns and proper nouns, as well as nouns related to addressing the audience. General nouns are more suitable to in-

16.078, df = 4306.2, p-value < 2.2e-16 ;ORG DE impromptu vs. SI EN DE (impromptu): t = -1.1415, df = 5559, p-value = 0.2537 - not significant

vestigate in a more detailed approach. Therefore, we look at the general nouns with the highest relative entropy in each subcorpus in more detail.

For English, this is *policy* for the interpreting subset and *enforcement* for originals. As *enforcement* (AvS = 13.2) occurs only in the originals subcorpus, instead, the noun with the second highest relative entropy, *legislation*, is investigated. For German, the only distinctive general noun *Sektor* (AvS = 10.97) for interpreting does not occur in the originals subcorpus. The most distinctive general noun for originals *Sicht* (AvS = 5.2) does not occur in the interpreting subcorpus. The general noun with the second highest relative entropy *Politik* is therefore selected for further inspection.

NOUN	AvS in ORG	AvS in SI
policy	*6.646364	*6.834648
legislation	*7.641000	*7.994444
Politik	*9.301026	*9.588000

Table 5.22: Average surprisal (AvS) for distinctive nouns by subcorpus

Table 5.22 shows that all the nouns investigated are more expected with lower AvS in the original subcorpus, regardless of the noun being distinctive for originals or interpreting in the first place. However, this trend is not significant, as indicated in the table by *[34].

Analysing potential triggers and translations of low surprisal nouns A parallel approach is used again to investigate simplification from the source to the target: potential triggers of low surprisal nouns in the target are investigated, and translations of low surprisal nouns of the source are analysed. The same approach as described for low surprisal verbs above is also applied for nouns: POS NOUN with surprisal ranging from 2.0 to 2.9 are extracted from the different subcorpora. The results with the 10 most frequent low surprisal nouns are given in Appendix A.8.

As a first general observation it can be stated that for the English subcorpora, low surprisal nouns are slightly more frequent in interpreting compared to originals (cf. Table 5.23). In German, opposite to the results for English nouns and for lexical verbs as described above, low surprisal nouns are more frequent in originals.

ORG EN		SI DE EN		ORG DE		SI EN DE	
rf	fpm	rf	fpm	rf	fpm	rf	fpm
326	5,021.49	307	5,651.38	822	14,797.48	721	13,397.50

Table 5.23: Frequencies for low surprisal nouns with spr = 2.0 - 2.9.

For English, the most frequent low surprisal noun, both in originals and in interpreting, is *people*. All occurrences of *people* with surprisal ranging from 2.0 to

[34] no significant difference: Wilcoxon rank sum test with continuity correction

2.9 are extracted, and their corresponding source or target expressions are manually aligned. For aligned items with formal equivalence, i.e. corresponding items that are also nouns, AvS is calculated. Other alignment categories chosen are *none* when no corresponding source or target expression, respectively, was found, and *other* if a source or target expression was found but belongs to a different POS (e.g. *people* - noun vs. *sie* - pronoun). The focus of the investigation was formal equivalence (translation also realised by a noun) and not semantic equivalence (e.g. *(British) people - Großbritannien* was seen as formally equivalent, although semantic differences are obvious). Table 5.24 gives an overview of low surprisal *people* in the source and the corresponding translations, as well as low surprisal *people* in interpreting and the corresponding triggers.

ORG EN (source *people*, spr = 2.++)	SI EN DE (target expressions)
1 x (nationality) people	10 x Volk, Volkes, Bevökerung, Wähler, Iren, Großbritannien (AvS = 10.61)
4 x people	4 x Menschen, Minderjähringe (AvS = 5.97)
1 x people	1 x none
2 x people	2 x other
ORG DE (source triggers)	**SI DE EN (target *people*, spr = 2.++)**
5 x Jugendlichen, (junge) Menschen, Jugend, Minderjährige (AvS =9.96)	5 x (young) people
3 x Menschen, Volk, Straftäter (AvS = 6.87)	3 x people
2 x hapax legomena noun (highest srp)	2 x people
1 x none	1 x people
2 x other	2 x people

Table 5.24: Low surprisal *people* and corresponding source/target expressions.

The following differences can be observed. Low surprisal *people* in the source refers mainly to different nationals, in the form of nationality adjective + noun (*Irish people, British people...*), which interpreters translate with a diverse set of target expressions. The literal translation with formal equivalence (e.g. *britischen Volk*) is only used 3 times. Nominative *irischen Volk* is particularly low in surprisal with spr = 0.94, whereas *irischen Volkes* in genitive is unexpected (spr = 14.29). This example shows how German's richer morphology with more inflected forms leads to large differences in surprisal, although it can be questioned whether processing differences of such extent can really be linked to the grammatical form of a word, or whether processing is not rather linked to the (semantic) meaning of a word. Basing the calculation of surprisal on the basis of lemma might be a way, especially for languages that are morphologically rich such as German, to better approximate cognitive processing.

Generally, it can be stated that there is no standard translation in German for low surprisal *people* of the English source, and average surprisal for these diverse target expressions is relatively high (AvS = 10.61). Therefore, no simplification from the source to the target can be observed in this respect. A more generic use of low surprisal *people* of the English source is translated with relatively low average surprisal target nouns (AvS = 5.97). Furthermore, translations with no formal equivalence (using a different word class, e.g. pronoun *Sie*) can also be observed, as well as instances where no translation was found. For these two categories, no surprisal was extracted.

Looking at source triggers of low surprisal *people* in the target, it can be observed that the most frequent use of this low surprisal noun is in the adjective-noun combination *young people*. Only once is the trigger the formally equivalent *junge Menschen*, but with far higher surprisal than the translation (srp = 9.41). All of these occurrences of low surprisal *people* in the target are triggered by a diverse list of source expressions with AvS that is higher than AvS for all nouns in the subset. Other uses of low surprisal *people* in the target are triggered by nouns that are more toward the average for the whole word class with AvS = 6.87[35]. Interestingly, hapax legomena in the source with the highest possible surprisal is translated twice with low surprisal *people*.

Looking at this direction alone, source triggers of low surprisal *people* in the target, would strongly point to simplification from high surprisal nouns in the source to a low surprisal noun in the target. However, in the opposite direction, looking at translations of low surprisal *people* in the source, also triggers a high surprisal target expression. Therefore, the trend observed seems to be linked to the English noun *people* and its translation entropy for German.

In German, the 10 most frequent low surprisal nouns (cf. Appendix A.8) mainly include nouns that are used to address the audience such as *Damen, Herren* or *Frau*. A more general low surprisal noun that is used in both ORG DE and SI EN DE is *Bereich*. Table 5.25 gives an overview of the translations of low surprisal *Bereich* in the original source, as well as the source triggers for low surprisal *Bereich* in interpreting.

Firstly, it can be observed that all the low surprisal and very expected occurrences of *Bereich* in originals and interpreting are in the context of the set phrase *in diesem Bereich*. Interestingly, this expected use of *Bereich* in the source is mostly left out in the target by the interpreters (in 80% of the cases). Only twice is this low surprisal input also rendered in the target speech; once with a phrasal expression *when it comes to*, which is also highly expected as a whole (AvS of this phrase = 3.02). The second translation of low surprisal *Bereich* is *here* (srp = 9.35).

Triggers of low surprisal *Bereich* in interpreting are more diverse phrases. If the head of the phrase is also a noun, this was seen as formally equivalent and AvS can be

[35] AvS for all nouns in ORG DE is 8.57

ORG DE (source *Bereich*, spr = 2.++)	SI DE EN (target expressions)
10 x	2 x other 8 x none
ORG EN (source triggers)	**SI EN DE (target *Bereich*, spr = 2.++)**
4 x noun (AvS = 2.05) 2 x phrasal 3 x none	9 x

Table 5.25: Low surprisal *Bereich* and corresponding source/target expressions.

compared. These source triggers are *in this area* and *across the board*, with equally low surprisal as *in diesem Bereich*. Furthermore, the two other phrasal source triggers *on this* and *doing that* cannot be compared directly, but are nonetheless slightly higher in average surprisal (AvS = 6.29). In 3 cases, a low surprisal *Bereich* in the target is added by the interpreter and does not have a source trigger. This is interesting, as the occurrences of phrasal *Bereich* do not convey much content, but seem to be used rather as a filler, e.g. where the interpreter may be stalling (cf. Section 2.2.2).

As low surprisal *Bereich* in originals is mostly left out in the target and at the same time low surprisal *Bereich* in interpreting has diverse source triggers, a slight tendency to simplification can be stated for this investigation.

Analysing potential triggers and translations of high surprisal nouns In a last step, the top of the surprisal range for nouns is investigated in order to see if the simplification hypothesis holds true from the source to the target for high surprisal nouns and to explore how very unexpected nouns of the target are triggered. Nouns with surprisal greater than 16 are extracted for English and greater than 15 for German[36]. Table 5.26 shows the raw frequencies (rf) and relative frequencies (fpm) of these high surprisal nouns for the different subcorpora and confirms the expected tendency of high surprisal nouns to be far less frequent in interpreting than original speeches in both languages studied.

ORG EN		SI DE EN		ORG DE		SI EN DE	
rf	fpm	rf	fpm	rf	fpm	rf	fpm
436	6,715.85	263	4,841.41	1,045	18,811.88	910	16,909.47

Table 5.26: Frequencies for high surprisal nouns with spr > 16 in English and srp > 15 in German.

As source and target items have to be aligned manually, a randomly selected subset of 50 high surprisal nouns in each subcorpus is selected. Categories chosen for the aligned items are, as for the low surprisal nouns above, *noun* for aligned items

[36]Surprisal for nouns range from 0 to 16.48 in English and from 0 to 15.88 in German

with formal equivalence, *hapax legomena noun* for aligned items that only occur once (no surprisal can be extracted), *none* where no aligned item can be found and *other* for alignments to other POS than NOUN (e.g. noun *consolidation* in source translated verbally with *konsolidieren* in interpreting). Furthermore, randomly selected high surprisal nouns include compound nouns in source and/or target, such as the following examples. The high surprisal noun *farmer* is the second part of the compound *dairy farmer*, which has two individual surprisal values (srp = 13.16 + 16.48). In German, compounds are spelt as one string. Therefore, surprisal of the German translation *Milchbäuerin* cannot be compared directly to the English equivalent. For such alignments, the category *compound* was created as surprisal for compounds as single words German cannot be compared to multi-word surprisal for compounds in English, where two or more words make up the compound. Depending on the fixedness or frequency of the compound, surprisal for the last noun in the compound is influenced by the other components of the compound noun.

Table 5.27 gives the result of the manual alignment for the 50 randomly selected high surprisal nouns. High surprisal nouns in the English source speeches are mostly also translated with nouns that are just below or at maximum surprisal for German nouns: 32% of nouns in the target are also high surprisal nouns, and further 22% of the target nouns are *hapax legomena nouns*, and therefore most unexpected. The high surprisal nouns used in German to translate high surprisal English nouns have an average surprisal of 14.04, which is higher than comparable triggers (AvS 13.19) and far above AvS for all comparable nouns in German (AvS = 8.81, cf. Table 5.28).

The same trend can be observed for high surprisal nouns in the German target and their triggers in the English source. High surprisal nouns in the target are mostly also triggered by high surprisal nouns in the source. With an AvS of 13.32, these nouns are higher in AvS than comparable target expressions (AvS = 12.36) and also above average surprisal for all nouns in English (AvS = 9.60, cf. Table 5.28). This shows that no general simplification trend can be observed for high surprisal nouns being translated from English into German. However, when dealing with English noun input that is very unexpected and therefore difficult to process cognitively, German interpreters tend to omit these nouns twice as often as they would add such unexpected content with no trigger in the source: 14% of high surprisal nouns from the source were left out in the target compared to only 6% of cases where a high surprisal noun in the target was added with no trigger in the source. Theoretically, this omission of content that is difficult to process could make the interpreting product more simple than the original it was based on. Examples 31 and 32 show the English source segment with high surprisal content and the aligned interpreters output where *vulnerability* was omitted. Only comparing the surprisal of the nouns for the two sentences would confirm such a simplification trend. With noun AvS of 11.97, the English source is cognitively harder to process than the German interpretation with noun AvS = 8.8. However, comparing surprisal for the entire segment, the English source – including the high surprisal noun – is easier to process on average (AvS =

ORG EN (source srp = 16.++)	SI EN DE (target expressions)
50 x	16 x (32%) noun, AvS = 14.04 (from those 9 cognate nouns) 11 x (22%) hapax legomena noun 7 x (14%) compound 7 x (14%) none 9 x (18%) other
ORG EN (source triggers)	SI EN DE (target srp = 15.++)
26 x (52%) noun, AvS = 13.32 (from those 11 cognate nouns) 5 x (10%) hapax legomena noun 13 x (26%) compound 3 x (6%) none 3 x (6%) other	50 x
ORG DE (source srp = 15.++)	SI DE EN (target expressions)
50 x	23 x (46%) noun, AvS = 12.36 (from those 11 cognate nouns) 3 x (6%) hapax legomena noun 7 x (14%) compound 5 x (10%) none 12 (24%) other
ORG DE (source triggers)	SI DE EN (target srp = 16.++)
23 x (46%) noun, AvS = 13.19 (from those 10 cognate nouns) 10 x (20%) hapax legomena noun 11 x (22%) compound 4 x (8%) none 2 x (4%) other	50 x

Table 5.27: High surprisal nouns and corresponding source/target expressions.

7,67) than the German interpretation (AvS = 8.62). Omitting a high surprisal word does not necessarily lead to lower surprisal output and, generally, simplification of high surprisal nouns cannot be observed for the translation direction English into German in none of the categories investigated.

(31) ... the (2.65) *vulnerability* (16.48) of (5.17) people (7.09) living (8.43) in (1.53) shacks (12.33) ...

(32) ... der (5.09) Personen (13.56) die (4.87) jetzt (8.27) Hütten (4.04) bauen (15.88) ...

For the translation direction German into English, the reverse tendency can be observed. High surprisal nouns of the source (Example 33) are mostly translated with nouns that are lower in surprisal (Example 34). Average surprisal for these target nouns is below average surprisal for comparable source triggers (AvS = 12.36 vs. AvS = 13.32). Although this is still above the average surprisal for comparable nouns in English (AvS = 9.17), it is well below the maximum surprisal for nouns in English (spr = 16.48). Furthermore, hapax legomena nouns are rarely used (6% of cases) in the target as a translation of high surprisal source nouns. This could be interpreted as simplification from the source to the target.

(33) ... dann sollten wir auch die polizeilichen *Ermittlungen* (15.30) abwarten ...

(34) ... let 's wait for the police *investigation* (5.39) ...

When investigating triggers of high surprisal nouns in SI DE EN it can be observed that they are mostly based on high surprising nouns (46% of cases) or hapax legomena nouns (20% of cases) in the source. Furthermore, it can be observed that the category *none* is rarely used in this translation direction, neither for target expressions nor for source triggers. This means that high surprisal nouns in interpreting are rarely added without a trigger in the source, and high surprisal nouns in the source are also rarely omitted in the target.

ORG EN	SI EN DE (target)	ORG DE (comparable)	all SI EN DE
16.++	AvS = 14.04	AvS = 13.19	AvS = 8.81
ORG DE	SI DE EN (target)	ORG EN (comparable)	all SI DE EN
15.++	AvS = 12.36	AvS = 13.32	AvS = 9.17
ORG EN (triggers)	SI EN DE	SI DE EN (comparable)	all ORG EN
AvS = 13.32	15.++	AvS = 12.36	AvS = 9.60
ORG DE (triggers)	SI DE EN	SI EN DE (comparable)	all ORG DE
AvS = 13.19	16.++	AvS = 14.04	AvS = 8.57

Table 5.28: High surprisal nouns, corresponding source/target nouns and average surprisal for comparable data.

Even though it is not possible to see a common trend in the categories described above for the translation or triggers of high surprisal nouns for both languages, one general observation in relation to cognitive processing and activation of translation equivalents can be made. When looking at the translation equivalent noun-noun pairs, it can be observed that around 20% of all studied high surprisal noun equivalents, which correspond to about half of the aligned noun-noun pairs, are cognate pairs. This is true for both translation directions and can be observed regardless of a high surprisal nouns in the source (Example 35 and 37) or used as a target expression

(Examples 40). Example 37 particularly shows the easy activation process for cognates as the high surprisal noun in the source *Entbürokratisierung* is first translated by a most unexpected noun (*debureaucratization*) and then verbally corrected with a more expected translation *get (10.02) rid (5.18) of (0.04) red (10.24) tape (0.74)* with much lower maximum and average surprisal (AvS = 5.24).

(35) ... you told us that the *flag* (16.48) and the anthem would be dropped ...

(36) ... Sie haben gesagt ja die *Flagge* (hapax legomena noun) wird gestrichen die Hymne wird gestrichen ...

(37) ... haben wir uns gewundert dass *Entbürokratisierung* (15.88) eine Wirtschaftskrise braucht um ...

(38) ... we tried hard to bring about the *debureaucratization* (hapax legomena noun) get rid of red tape ...

(39) ... the *slogan* (16.48) British workers for British jobs must mean ...

(40) ... dieser *Slogan* (15.30) britische Arbeitnehmer für britische Arbeitsplätze muss bedeuten dass ...

Cognate activation is based on the principle of language non-selective lexical access and a multiplicity of cross-level activations. This means that when hearing the English word *Flag*, the German cognate *Flagge* is activated instead of *Fahne* or another word with similar meaning but different form. Therefore, cognates are seen to be cognitively easier to retrieve (Dijkstra et al., 2015). Studies show that interpreters opt for a cognate translation more frequently than translators (Oster, 2017). When doing so, they are faster in producing their translation as indicated by translation latency being generally shorter than for non-cognate translations (Defrancq, 2015). Furthermore, translation accuracy increases when using cognate translations (Chmiel et al., 2021). Using cognates for highly unexpected nouns can therefore be seen as a producer-oriented mechanism to facilitate cognitive load, which is shown here to be more prominent than the activation of a lower surprisal noun. Interpreters can be seen as making their lives easier by choosing the closest translation pair and thereby reducing their cognitive load. However, this producer-oriented mechanism does not result in a simplification of the interpreting product.

To summarise the finding for nouns, it can be said that analysing nouns and their expectedness in contexts is particularly interesting as this POS is distinctive for originals and not for interpreting. Differences can be seen for both languages studied. For

English, using a comparable approach and studying interpreting with English parallel data, the simplification hypothesis can be confirmed. Nouns in English interpreting are lower in surprisal than comparable originals and this trend is also confirmed when taking delivery mode into account. Low surprisal nouns are more frequent in English interpreting, whereas high surprisal nouns are more frequent in English originals. Furthermore, when looking at simplification from the source to the target, a simplification trend can also be seen for low and high surprisal nouns in English. For German, on the other hand, the opposite trend can be observed for all but one investigated category (high surprisal nouns are less frequent in interpreting than German originals); thus the simplification hypothesis cannot be confirmed when considering nouns in German.

5.2.5 Summary of lexical simplification via surprisal

The method of surprisal found a general tendency for simplification in simultaneous interpreting. Overall, interpreters were found to use more expected lexis and therefore more simple lexis than original speakers. This means that the interpreter as the recipient of original speech needs to process more unexpected words than the recipient of interpreted speech. By producing lower surprisal output, the interpreter facilitates processing for the listener. Whether this type of simplification is audience design (intended by the interpreter to reduce processing costs for the listener) or resulting from their own processing constraints cannot be answered by means of this analysis. However, the use of cognates for high surprisal nouns, also found in this study, points to more speaker-oriented, rather than producer-oriented mechanisms.

This general trend cannot be confirmed for all categories investigated in this study. Table 5.29 summarises the individual results for each subcategory. A general simplification trend can be observed when considering the entire speeches with all tokens, with AvS lower in interpreting than in comparable original speeches. For both verbal categories AUX and VERB a general trend of simplification can also be observed. This is true for AUX as a closed word class, where interpreters and original speakers use the same limited list of lexical items, but the context in which those auxiliaries are used is more standardised in interpreting. For VERB, the lexical items used can differ, but interpreters still use lower surprisal verbs. An exception to this trend is the investigation by delivery mode impromptu, where no simplification in interpreting was observed for auxiliary verbs or lexical verbs in German. The investigated category NOUN, on the other hand, showed a mixed picture in which simplification was observed only for English, but not for German in the comparable and parallel analysis. This is an interesting finding, as simplification is found to occur rather in verbal POS, which were found distinctive for interpreting when compared to originals (cf. Chapter 4), and not to the same extend in nominal POS, distinctive for originals when compared to interpreting. These two results, simplification more in verbal than nominal POS and in English more than in German, may be linked, as English tends

Category	Subcategory	Simplification hypothesis
ALL	all tokens	confirmed
	only impromptu	confirmed
AUX	all AUXs	confirmed
	only impromptu	confirmed for EN
		not confirmed for DE
VERB	all VERBs	confirmed
	only impromptu	confirmed for EN
		not confirmed for DE
	KLD results	partially confirmed
	parallel low surprisal	confirmed
	parallel high surprisal	confirmed
NOUN	all NOUNs	confirmed for EN
		not confirmed for DE
	only impromptu	confirmed for EN
		not confirmed for DE
	KLD results	not confirmed
	parallel low surprisal	confirmed for DE to EN
		not confirmed for EN to DE
	parallel high surprisal	confirmed for DE to EN
		not confirmed for EN to DE

Table 5.29: Simplification hypothesis confirmed/not confirmed for POS categories analysed.

to be more verbal than German. However, it could also be influenced by German's richer morphology and the resulting effects on string-based surprisal, as discussed above.

Furthermore, the comparative analysis showed that delivery mode, or preparedness of original speeches also influences simplification patterns. Lower surprisal was generally found in impromptu originals rather than in prepared and read source speeches. This can be seen as spontaneous production using more standardised patterns and less variation than a pre-scripted and read-out speech. Interpreting, in this sense, can be seen as more spoken than spoken, in line with this trend observed in the literature (Shlesinger and Ordan, 2012).

A more detailed investigation of high and low surprisal verbs and nouns revealed further interesting results. While low surprisal verbs were generally more frequent in interpreting than in originals in both languages, the same trend was observed only for low surprisal English nouns, but not for low surprisal German nouns. In line with this observation, high surprisal verbs as well as nouns were found to be used more frequently in originals than in interpreting in both languages.

The **RQ 2a** "Is simultaneous interpreting more simplified at the lexical level than

comparable originals in the same language?" can therefore generally be confirmed. Simultaneous interpreting is more simplified at the lexical level than comparable originals. It was confirmed for English interpreting from two different source languages. However, the target language has an influence, as the trend is more pronounced in English than in German.

The analysis also attempted to investigate parallel data and find triggers for simplification in interpreted speech. For **RQ 2b** on the triggers of lexical simplification, the following can be stated. For the lower range of surprisal, simplification was observed for verbs and English nouns when low surprisal output in interpreting was triggered by low surprisal input, added as a low surprisal filler or low surprisal pragmatic discourse marker, and particularly by using a standardised translation solution for a diverse range of input. This addition of low information content may be linked to the interpreting strategy of stalling (Kalina, 1998; Seeber, 2011; Liontou, 2011). Investigating the higher range of surprisal, it was observed that simplification occurred for lexical verbs in both languages studied. High surprisal input was translated with high surprisal output or occasionally left out. The same trend was observed for English nouns, but not for German. Simplification was occasionally found as the result of omitting highly unexpected input. On the other hand, high surprisal output was never added, but instead was always triggered by high surprisal input or hapax legomena words. This could be interpreted that for unexpected input, interpreters are not able to resort to standard material and use other means to deal with high processing costs, such as omission and the use of cognates (leading to high surprisal nouns in the target speech).

However, the small dataset used for modelling surprisal as an approximation of cognitive processing was found to be a shortcoming of the methodology applied. Hapax legomena occur relatively frequently in the dataset, especially in the German dataset with its rich morphology. In English 2.13% of all tokens (or 3619 tokens) occur only once; in German, this is true for 6.28% of tokens (or 6865 tokens). In comparable datasets such as the Europarl UdS only 0.1% of the English data and 0.8% of the German data occur as unique strings[37]. The small EPIC UdS dataset is not only problematic because a relatively high proportion of the data, all unique strings, are treated as very expected although, theoretically, they should be most unexpected in context and have the highest surprisal allocated. Other low frequency word forms might be over- or under-represented, as well as topics of speeches covered in the data may influence surprisal of lexical words.

Overall, it can be stated that the EPIC UdS dataset might be too small to represent general linguistic knowledge that is crucial for cognitive processing of words in context. With only 137 original speeches in English and 162 original speeches

[37]Europarl UdS EN: 19 039 out of 19 313 205 or 0,09858%; Europarl UdS DE: 97 341 out of 12,949,278 or 0,7517%

in German covering 52 individual speakers in English and 89 speakers for German, as well as 10 different topics covered in the dataset, the dataset is also fairly small to represent context knowledge of this particular setting. In order to get a more reliable picture of processing difficulties or ease of individual tokens in EPIC UdS, using a larger comparable dataset for modelling surprisal could be considered. For English, this could mean using English data from other EPIC corpora such as EPIC (Russo et al., 2012), EPICG (Defrancq, 2015), PINC (Chmiel et al., 2021) or ESIC (Macháček et al., 2021) for modelling spoken English in the EP. Unfortunately, the only fully comparable dataset available for German at the time of writing this thesis is ESIC. However, most of the other datasets within the EPIC family are similar in size or smaller compared to EPIC UdS. Therefore, even combining these resources to get a more comprehensive picture of linguistic and context knowledge of this particular spoken register might still be too limited. Another solution could be using a modified version of Europarl UdS, the written equivalent of the corpus data at hand, for calculating the language model as basis for EPIC UdS, or even a large general language model for English and German to better approximate overall linguistic and world knowledge that is used for language processing.

On a more general note, it should be discussed whether the 3 token context is enough to reflect cognitive processing, when even ear-voice span in interpreting is generally longer than 3 tokens (Defrancq, 2015). However, a larger context to be used for modelling surprisal is not feasible for such small corpora such as EPIC UdS and the modelling approach taken.

Furthermore, surprisal for the corpora investigated is modelled on word level as described in Section 5.2.1. However, language processing does not seem to be uniquely linked to the word forms used but rather concepts linked to these lemmas. Especially in morphologically rich languages such as German, differences in surprisal for the same lemma have a large effect, particularly within small LMs. For example, surprisal for the nominative noun *Volk* (EN: people) can be very low (*würde dem britischen Volk (spr – 0.93)* or *hat dem irischen Volk (spr – 0.94)*), while the same noun in genitive is highly unexpected, with surprisal ranging from 9.38 to 14.29 (*Entscheidung des irischen Volkes* (srp = 14.29)). Although nominative is more frequent in German overall and therefore should be easier to be processed, such large differences in surprisal do not seem to reflect actual processing differences. A solution to overcome this shortcoming would be to model surprisal as not string-based but rather lemma-based.

Despite these shortcomings of the approach used, the results of the analysis with simplification in interpreting observed in most of the investigated categories are very convincing. Moreover, Kunilovskaya et al. (2023b) use the same modelling approach with the same corpus data and a balanced dataset (resulting in even fewer data for language modelling), and also obtain convincing results.

Chapter 6

Exploring syntactic simplification

In addition to the differences in the lexical use of interpreting compared to originals (Chapter 5), the results of the exploratory analysis via relative entropy (Chapter 4) also point to syntactic differences. Interpreting is found to be verbal with clausal elements, whereas original speech was described to use more nominal parts-of-speech and prepositions, pointing to different syntactic structures using nominal and prepositional phrases. The following chapter investigates syntactic differences and links them to the research questions on syntactic simplification. Specifically, the following research questions on syntactic simplification are addressed.

RQ 3a: Can syntactic simplification measured via sentence length (comparable and parallel approach) and alignment analysis (parallel approach) be observed in interpreting?

Furthermore, considering the aspects for and against potential dependency length minimisation, the following research questions are asked to give a more detailed insight about syntactic complexity.

RQ 3b: Can syntactic simplification measured via dependency length minimisation be observed in the interpreting product (comparable approach, comparing originals and interpreting in the same language)?

RQ 3c: What are the triggers of a potential minimisation effect? Does shining-through of source language structure counterbalance a possible minimisation effect (parallel approach, comparing source and target speech)?

Sections 6.2.1 and 6.2.2 generally investigate the sentence level, comparing sentence length on word level for the comparable and parallel interpreting data, as well as investigating whether interpreters maintain the source speech sentence structure or rather split or merge source sentences in their output (**RQ 3a**). **RQ 3b** and **RQ 3c** focus on dependency length measures (Section 6.2.3), investigating syntactic simplification in a comparable approach, comparing originals and interpreting in the same language (**RQ 3b**) and then looking into the triggers of a potential minimisation effect (**RQ 3c**).

Generally, syntactic simplification has been defined in Section 2.5.2 as using shorter sentences and shorter length of dependency relations.

6.1 Investigating syntactic complexity: method

This chapter investigates syntactic simplification from various angles. A first general approach compares sentence length as defined in Section 3.1.2 for the German and English comparable dataset. This includes interpreted English from two different source languages (German and Spanish). Furthermore, a parallel approach compares sentence length in interpreting to sentence length in the original speeches on which the interpretation was based. This combined approach gives a first insight into syntactic simplification as it uses a general simplistic approach assuming that shorter sentences are simpler than longer sentences. A trend of the sentence length in interpreting to be shorter than in comparable original speech was also found in the literature (Bernardini et al., 2016; Liu et al., 2023).

Although simultaneous interpreting is generally a linear process (cf. Section 2.2.1), interpreters are assumed to not always maintain the syntactic structure of the source, such as splitting complex main clauses into several independent main clauses (He et al., 2016; Gast and Borges, 2023). In a second step, this thesis takes a parallel approach and looks at the sentence aligned dataset to investigate if interpreters split, merge, or maintain sentences of source speech. The tendency to split source input could be interpreted as simplification from the source to the target (Baker, 1996). Although this has been discussed previously in the literature in the context of interpreting strategies (Kalina, 1998) and other corpus-based studies (He et al., 2016; Gast and Borges, 2023), this thesis is the first study to our knowledge that investigates large-scale source-to-target aligned interpreting data to study how interpreters deal with source sentences in their target output. For this purpose, 1:1 sentence alignments, 1:n and n:1 alignments, as well as sentences of source or target, where no alignment was found are extracted from the aligned dataset described in Section 3.1.3 and quantitatively analysed. If a trend of more 1:n alignments is found, this could be an indication of simplification. In the opposite case (more n:1 alignments), this would point to densification rather than simplification. The trends discovered in the quantitative analysis will be verified in a qualitative study. To exclude an effect of translation direction, this investigation is conducted for English source speeches and their interpretation into German as well as German source speeches and their interpretation into English.

The results of Sections 6.2.1 and 6.2.2 are used to answer **RQ 3a**, which aims to explore whether syntactic simplification in interpreting can be observed comparing sentence length (comparable and parallel approach) and source-to-target sentence alignment data (parallel approach).

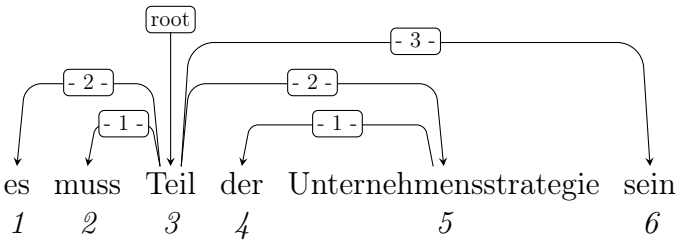


Figure 6.1: Dependency annotated example sentence.

The main focus of the syntactic investigations on simplification is to measure syntactic complexity and approximating syntactic simplification, via Dependency Locality Theory (DLT), and investigate a potential tendency of interpreters to minimise dependency distances (cf. Section 2.3.2). DLT assigns processing costs of all elements in a sentence via storage and integration costs (Gibson, 2000). Structures with longer dependencies are assumed to be more difficult to process Gibson (1998, 2000) and natural languages generally show a preference for shorter dependencies (dependency length minimisation, DLM) (Gildea and Temperley, 2010; Temperley, 2007). Although this generally applies to language comprehension and language production (Gibson, 1998, 2000; Diessel, 2005; Liu, 2008; Hawkins, 1995, 2004; Temperley, 2007, 2008; Liu et al., 2017; Ferreira and Dell, 2000; Arnold et al., 2000), in spoken language the focus seems to be more on facilitation of production (Rajkumar et al., 2016).

The study on syntactic simplification investigated via a dependency length analysis uses the Stanza parsed corpus version EPIC UdS V3 (cf. Section 3.1.4). Dependency length (DL) is used as a binary, asymmetric relation between two linguistic units (head and dependent) (Liu, 2008; Liang et al., 2017). Each *dependent* word in a sentence is linked to a *head*. Figure 6.1 illustrates this in an example, showing the dependency relations for a German example sentence. The arcs above the words extend from the *head* towards the *dependent* (Liu, 2008). The numbers below the words refer to the position of the word in the sentence and are used to calculate the dependency length (DL) or dependency distance (DD), which is indicated by the numbers above the arcs.

For calculation purposes[1], the sentence is seen as a string of words W_1 ... W_i ... W_n. DL between a head word W_a and dependent word W_b can be calculated as the difference of W_a and W_b. This results in a positive number (stating that the dependent word precedes the head) or a negative number (stating that the dependent word follows the head). Dependency length is always indicated as the absolute positive value.

In addition to looking at individual dependency length for each word, average dependency length (av DL) for each sentence is calculated (Equation 6.1). With "n" as the number of words in a sentence. "i" refers to the i-th word in the sentence. The root of a sentence generally does not have a head, and therefore DL for the root word is always defined as zero. End of sentence punctuation was inserted for parsing

[1]calculations are based on Liu (2008)

of the data but not included in the calculation of sentence complexity measures. For the example in Figure 6.1, the series of DL obtained by Equation 6.1 is 2 1 0 2 1 3. This results in av DL for the example sentence of $9/6 = 1.5$.

$$avDL(ofsentence) = \frac{1}{n-1}\sum_{i=1}^{n-1}|DL_i| \tag{6.1}$$

Further measures on syntactic complexity used in the investigation of syntactic simplification via DLT are maximum dependency length (max DL) of a sentence and the percentage of adjacent dependencies, with DL = 1. For the example in Figure 6.1, maximum dependency length is the distance between the dependent *sein* (word number 6) and its head *Teil* which results in a maximum distance of 3. The number of adjacent dependencies (DL = 1) can be counted as 2 out of 6 (or 33,3%). Lower values for average dependency length and maximum dependency length are defined as lower complexity of a sentence, whereas the higher the number of adjacent dependencies, the less complex a sentence. The syntactic complexity indicators are investigated for all subcorpora, as well as separately for short and long sentences. Besides these quantitative analyses using a comparable approach, Dependency Locality Theory is also used to investigate individual dependency relations with regard to their frequency and average length in interpreting and comparable originals. Results of these analyses are investigated qualitatively, looking into potential reasons for observed differences by comparing source and target sentences for individual dependency relations. As DLM is generally found to a much higher degree in English than in German (Gildea and Temperley, 2010; Futrell et al., 2015; Dyer, 2023), it will be interesting to see if this holds for interpreting in these languages.

The quantitative analyses of Section 6.2.3 aim to answer **RQ 3b** on whether syntactic simplification in interpreting can be measured via dependency length minimisation when interpreting is compared to original speeches in the same language. The qualitative analyses based on results of the quantitative findings for individual dependency relations, their frequency and average length aims to answer **RQ 3c** with the focus on triggers of a potential minimisation effect and the question whether shining-through of source language structure counterbalances a possible minimisation effect.

6.2 Syntactic analysis

6.2.1 Sentence length

As a first general approach to assess syntactic complexity, the sentence length for the different subcorpora is compared. The Spanish original dataset is also included to explore potential factors influencing the sentence length in the target speech. Particular long or short sentences in the source might have an effect on sentence length

and complexity in the target language. However, the main focus of the study is a comparable approach, investigating the comparable English and German subcorpora.

subcorpus	N	mean	SD	median	IQR	range
ORG EN	3622	19.16731	12.46648	16	15	97
SI DE EN	3661	16.75116	10.43924	14	11	122
SI ES EN	3076	18.24122	11.76522	16	14	107
ORG DE	3409	17.41302	11.684734	15	14	100
SI EN DE	4076	14.87341	8.905429	13	11	71
ORG ES	2537	22.02601	17.49331	18	20	156

Table 6.1: Sentence length overview.

From a comparable perspective, comparing interpreted speeches in one language with original speeches in the same language, we can observe that interpreters produce significantly shorter sentences than original speakers (cf. Table 6.1). English originals are significantly longer than interpreting into English from both source languages, German and Spanish[2]. This trend is stronger for interpreting with German as source language (SI DE EN) than for interpreting based on Spanish (SI ES EN), which can be linked to the particular long sentences in the Spanish source (mean sentence length ORG ES = 22.03 words vs. ORG DE = 17.41). The same trend can also be observed in German with interpreting into German (SI EN DE) being significantly shorter than comparable German original speeches[3].

Figure 6.2 illustrates that interpreted speeches are generally shorter in sentence length than comparable originals. This is not only true for mean and median sentence length (*mean* visualised as a diamond in the boxplots, *median* visualised as centre line in boxplots). Besides more outliers with very long sentences in interpretations into English (dots in the boxplots), all the quartiles displayed for interpretations are lower than for comparable originals (boxes and upper whiskers). This is true for comparable English with interpreting from both source languages, as well as for the German comparable dataset.

From a parallel perspective, it can be observed that interpreters produce shorter sentences than the original speeches on which they were based. SI EN DE is significantly shorter than ORG EN, SI DE EN is shorter than ORG DE, and SI ES EN is shorter than the source speeches ORG ES[4].

[2]Welch Two Sample t-test for parametric data: ORG EN vs. SI DE EN: t = 8.9625, df = 7037.1, p-value < 2.2e-16; ORG EN vs. SI ES EN: t = 3.1235, df = 6622, p-value = 0.001795

[3]Welch Two Sample t-test for parametric data: ORG DE vs. SI EN DE: t = 10.411, df = 6283.4, p-value < 2.2e-16

[4]Welch Two Sample t-test for parametric data: ORG EN vs. SI EN DE: t = 17.194, df = 6467.7, p-value < 2.2e-16; ORG DE vs. SI DE EN: t = 2.5049, df = 6838.8, p-value = 0.01227; ORG ES vs. SI ES DE: t = 9.3, df = 4288.9, p-value < 2.2e-16

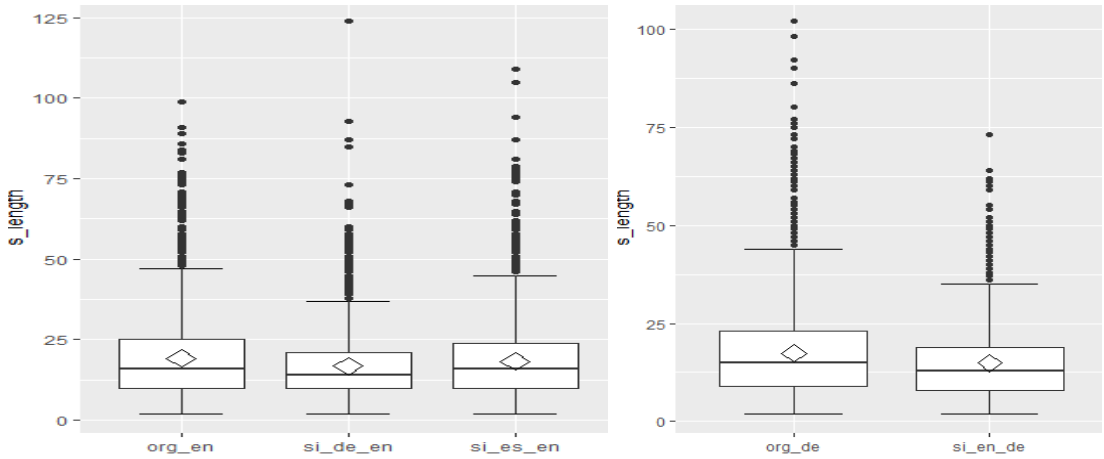


Figure 6.2: Sentence length for English (left) and German (right) comparable data.

Long sentences in the Spanish source speeches have an effect on target production, with longer sentences in English interpreting based on Spanish (SI ES EN) compared to English interpreting with German as source (SI DE EN).

The parallel perspective also allows one to get a first general insight into the causes of shorter sentences in interpreting. Interpreters do not just shorten sentences, but rather break up (longer) sentences in the source into several shorter sentences in the target output (Table 6.2). This is indicated by a higher number of sentences produced by original speakers compared to the number of sentences produced by interpreters for these source speeches.

	no. of source sentences	no. of target sentences
ORG EN > SI EN DE	3696	4173
ORG DE > SI DE EN	3409	3661
ORG ES > SI ES EN	2537	3076

Table 6.2: Number of sentences sources vs. target.

To summarise the sentence length analysis, simplification on the syntactic level can be observed through shorter sentences in the comparable and parallel corpus analysis. This is in line with trends observed in the literature (Bernardini et al., 2016; Liu et al., 2023).

6.2.2 Sentence alignment

The previous section shows that interpreters produce more, but shorter sentences than the original speakers on which the interpretation was based. In order to get a more comprehensive picture of the underlying process, this section investigates the sentence aligned corpus (cf. Section 3.1.3). This investigation focusses on the language pair English and German, with English original speeches interpreted into German and

German original speeches interpreted into English. All possible alignment pairs were grouped into more general alignment types: *equivalent* for 1:1 and 2:2 alignments, *split* for 1:n alignments, *merge* for n:1 alignments and *none* for 1:0 and 0:1 alignments. Tables 6.3 and 6.4 give an overview of source-to-target sentence alignment for English to German and German to English interpreting respectively. It includes the entire EPIC-UdS dataset for this language combination.

	no. of aligned pairs	percentage	alignment type	no.	percentage
1:1	2081	60.32%	equivalent	2141	62.06%
2:2	60	1.74%			
1:2	555	16.09%	split	716	20.75%
1:3	114	3.30%			
1:4	40	1.16%			
1:5	6	0.17%			
1:7	1	0.03%			
2:1	211	6.12%	merge	238	6.90%
3:1	21	0.61%			
4:1	6	0.17%			
1:0	269	7.80%	none	355	10.29%
0:1	86	2.49%			
total	3450	100%	total	3450	100%

Table 6.3: Sentence alignment for ORG EN to SI EN DE.

	no. of aligned pairs	percentage	alignment type	no.	percentage
1:1	2110	64.96%	equivalent	2148	66.13%
2:2	43	1.33%			
1:2	404	12.45%	split	533	16.45%
1:3	98	3.02%			
1:4	20	0.62%			
1:5	10	0.31%			
1:7	1	0.03%			
2:1	173	5.33%	merge	196	6.04%
3:1	22	0.68%			
4:1	1	0.03%			
1:0	280	8.64%	none	362	11.17%
0:1	82	2.53%			
total	3245	100%	total	3245	100%

Table 6.4: Sentence alignment for ORG DE to SI DE EN.

In both language combinations, by far the largest share of the original sentences is translated with a one sentence equivalent as in the examples in Table 6.5. 1:1

alignments make up 60.32% of alignments for original English interpreted into German and 64.96% of alignments for original German interpreted into English. The share slightly increases if 2:2 alignments are included in this category (62.06% and 66.13% respectively). The latter are two aligned source sentences with two aligned target sentences such as in example 2:2 in Table 6.5, where the first target sentence contains is based on two sentences of the source content of the first (source: *Getriebene + reagieren* – target: *pushed into a situation + react*).

This shows that maintaining the general macrostructure of the speech by keeping sentence equivalence is by far interpreters' preferred production type. It can be assumed that by maintaining the source text sentences structure, no additional cognitive resources have to invested when the source structure can be imitated.

	ORG DE	SI DE EN
1:1	das ist ein perverses System ↲	and that's a perverse system ↲
1:1	und das gehört abgeschafft und verboten ↲	and that's got to be got rid off / and banned ↲
2:2	wir sind Getriebene ↲	we're pushed into a situation and have to react to it ↲
	und wir reagieren nach meinem Dafürhalten zu spät ↲	and I think we're doing that too late ↲

Table 6.5: Equivalent sentence alignment for ORG DE to SI DE EN.

The second largest category are sentences of the source speech that are spilt into several sentences (1:n). This is true for both interpreting directions and even more pronounced for English originals that are interpreted into German than for the opposite direction (20.75% vs. 16.45%). Most of these sentence splits in the target are 1:2 splits, but splits of up to 1:7 were found in the corpus data[5].

Table 6.6 gives examples for such sentence splitting for German original speeches that are interpreted into English. In the 1:2 split sentence example, the source structure is very similar to the structure of the first target sentence. The subject in the source *die grausame und alltagliche hausliche familiäre und lokale Gewalt gegen Frauen und Mädchen* matches the subject in the target *terrible daily family violence against women and girls* with the omission of *häuslich* and *lokale*. The subject is followed in both sentences by a copular verb. The complement in the German source sentence is with 11 words much longer than the complement in the target rendition: *das brennendste schlimmste und von unseren Medien am meisten ignorierte Problem* vs. *the worst thing*, which can be seen as a summary translation. The simultaneous interpreted rendition splits the three premodifications of *Problem* into a short complement of three words and starts a new sentence for the dense modification *von*

[5]The Trados Studio 2021 alignment tool only allows for 1:n and n:1 alignments with $n \geq 0$ and $n \leq 5$. The cases of $n = 7$ were extracted manually.

	ORG DE	SI DE EN
1:2	die grausame / und alltägliche häusliche familiäre / und lokale Gewalt gegen Frauen und Mädchen / ist für mich ohnehin das brennendste schlimmste / und von unseren Medien am meisten ignorierte Problem ↲	terrible / daily / family / violence against women and girls / is the worst thing / from my point of view though ↲
		and yet our media don't tend to give it much attention ↲
1:4	die Kommissarin / hat im Nachgang der Paraphierung des Abkommens / mit den Vereinigten Staaten erreicht dass ein wesentliches Anliegen nämlich dass europäische Beamte in Washington vor Ort sind / die Extrahierung der Daten überwachen und in Missbrauchsfällen / gegebenenfalls auch blockieren kann / dass das / Einzug in das Abkommen gefunden hat namentlich in Artikel zwölf eins ↲	and the Commissioner a/ afterwards were in/ in approving and streamlining the agreement did achieve / euh an important concern / even frightened that the European Official would be in Washington on the spot ↲
		and it/ it would monitor the extraction of data ↲
		and t/ t/ and to block extraction of data if there is abuse ↲
		euh and that's now incorporated into a/ article twelve / euh one of the agreement ↲

Table 6.6: Split sentence alignment (1:2 and 1:4) for ORG DE to SI DE EN.

unseren Medien am meisten ignorierte. The premodifying adjective *brennendste* in the source is omitted in the target rendition.

Omission and restructuring do not make the target rendition overall shorter (28 words vs. $18 + 12 = 30$ words). However, the split premodification of the complement noun makes the interpreted rendition less dense.

The 1:4 sentence example (cf. Table 6.6) shows a long, multilayer embedded clause in the source that is split into several independent sentences in the target. In doing so, the interpreter reduces embeddedness and therefore increases simplicity in the target when compared to the source: The source sentence includes several subordinate clauses that are further postmodified by additional subordinate clauses

and a non-clausal modification *(... (dass.... Anliegen (nämlich dass... blockieren kann) (dass das ... gefunden hat))* and *(namentlich in ...)*. The maximum level of embeddedness in the target is one subordinate clause *(...(that the ....)* – first target sentence) and *(...(if....))* – third target sentence).

	ORG EN	SI EN DE
1:2	at the outset / let me put on the official record of the house / my thanks to all of my colleagues / both those who support / and those who oppose / the proposal that I'm putting forward / for their contributions for their input and in particular for their helpful advice and guidance along the road ↲	ich möchte / zuerst einmal / allen Kollegen danken auch für das Protokoll danken / die meine Vorschläge unterstützt haben ↲
		ich möchte ihnen danken für ihr Beiträge aber vor allem auch für ihre Ratschläge ↲

Table 6.7: 1:2 sentence alignment for ORG EN to SI EN DE.

The same trend can be described for the direction original English interpreted into German (cf. Tables 6.7 and 6.8). In the 1:2 example (Table 6.7), the source sentence consists of 53 words whereas the two target sentences only use 17 and 14 words respectively. With 31 words in total, the two target sentences combined are notably shorter than the source they are based on. However, this is not the driver of simplification in this case. The syntactic structure used in the long source sentence is much more complex than in the two shorter target sentences. The source sentence uses an elliptical structure where subject and predicate are omitted (e.g. *(I would like to give) my thanks to* followed by multiple embedded clauses and three prepositional objects. The first embedding (*those who support and those who oppose the proposal*) contains a further embedded clause (*that I'm putting forward*). This second embedded clause is left implicit in the interpreted speech and the second part of the first embedded clause (*and those who oppose*) is omitted completely. Furthermore, the original speaker uses four nouns to express what they are thanking for: *contributions* and *input*, *advice* and *guidance*. The noun pairs can be seen as synonymous. The interpreter avoids repetition and only gives a translation for one of the respective nouns (*Beiträge* and *Ratschläge*). Furthermore, modifications of these nouns that are present in the source are left out in the target (*helpful, along the road*), as well as *at the outset* at the beginning of the source sentence. These omissions render the target text less dense. The interpreter's repetition on structural level for both target sentences (*ich möchte (...) danken (...) für*) on the other hand makes the target speech simpler overall.

We interpret this as simplification from the source to the target for this example, that is achieved by (1) shorter and less complex sentence, (2) repetition of sentence structure in both target sentences, (3) omission of repetitive content words and modifications.

	ORG EN	SI EN DE
1:7	but the second point with regard to what this directive is doing / and possibly the most important of all / is that for the first time it recognizes the contribution of session musicians by establishing a fund to allow them to ensure / that they have a return and remuneration for their work / which has been exploited by people / over a long period of time for which they may only get a one-off payment / if they're lucky enough to get that ↲	aber ein zweiter Punkt ist wichtig ↲
		und der ist vielleicht am wichtigsten ↲
		zum ersten Mal wird festgestellt dass es auch Studiomusiker gibt ↲
		und diese werden nun in ihrem Wert aufgewertet ↲
		sie werden F/ einen Fond haben der von Personen gemanagt wird ↲
		und sie werden eine Zahlung bekommen ↲
		also in der Vergangenheit hatten sie nur ein Mal das Glück überhaupt Zahlungen aus diesem Fond zu bekommen ↲

Table 6.8: 1:7 sentence alignment for ORG EN to SI EN DE.

The 1:7 example (cf. Table 6.8) is a long and complex source sentence with multiple embedded structures being split into 7 separate main clauses in the target, out of which only one contains an embedded structure (subordinate *dass es auch Studiomusiker gibt* in the third sentence). Moreover, the subject of the source sentence is very long and complex: *the second point with regard to what this directive is doing and possibly the most important of all.* The interpreter omits the postmodification of the head noun *point – with regard to what this directive is doing –* and leaves it

implicit in the target. This information can be implied from the context/previous utterances. Furthermore, the interpreter splits the subject's postmodifications and produces two short sentences for this instead. The main verb in the source is *is*, a copular verb, followed by a complement clause that includes multilayered embedded structures: *...(that... (by establishing... (to ensure that ...(which has...)(for which... (if ...)))))*. These individual layers of embedded structures are split by the interpreter to produce separate individual sentences.

This type of sentence splitting simplifies the interpreting product compared to the source sentence on which it was based.

	ORG DE	SI DE EN
1:0	nun fö/ euh fährt die / euh österreichische Eisenbahn / vier euh fünf Mal am Tag diese Strecke / ↲	–
1:0	das tut mir aufrichtig leid ↲	–
1:0	vielen Dank ↲	–
0:1	–	thank you ↲
0:1	–	that's my first point ↲
0:1	–	this is what we need to do ↲

Table 6.9: *None*-sentence alignment for ORG DE to SI DE EN.

The third largest alignment category (cf. Tables 6.3 and 6.4) is the group of *none*: source or target sentences where no target or source equivalent was found respectively (10.29% for English to German and 11.17% for the German to English alignments). The majority of *none*-alignments are 1:0 alignments, where the original speaker produced a sentence which was completely omitted by the interpreter. This is the case for 7.8% of alignments for German to English and 8.64% for English to German. A more detailed investigation of the type of source sentences or sentence equivalents that are left out shows that such omissions often occur at the end of the speech where the source speaker typically ends with *Thank you*, *Danke* or *Vielen Dank* (cf. example in Table 6.9). The particular interpreting setting in the European Parliament with very short speaker interventions and many different languages spoken may be the main reason for skipping the typical *Thank you* at the end of a speech. The interpreter already lags behind the original speaker; however, knows about the importance of quickly having to free the interpreting channel so that a booth colleague can take over. It is also possible that the interpreter has to select a new audio input channel in the correct language in time. For one or the other reason, there is a need not to lag behind the original speaker for too long at the end of a speech, and therefore interpreters may omit standard material that does not add to the point of discussion.

Besides these omitted sentence equivalents at the end of the speeches, this category also includes sentence omissions within the speech, where the content of the original speech is left out, such as in the first and second example in Table 6.9. In the

first example, the interpreter leaves out additional details regarding the matter of discussion (the previous sentence talks about the Austrian railway and the omitted sentence would provide detail about the frequency of railway service). In the second example, the interpreter omits a source sentence which conveys the speaker's attitude towards the matter being discussed.

These type of omissions (standard material, examples or speaker's attitude) generally should not lead to coherence problems in the target and therefore can be assumed to render the interpreter's output less dense and more simple than the original speech it was based on.

Additions to the target speech - the category of 0:1-alignments - are much less frequent. They may be typical end of speech sentence additions (*thank you*), discourse structuring inserts (*that's my first point* - where the original only mentions *und es ist zweitens...* and the interpreter adds the correct numbering of items) or making the speakers intention, as interpreted by the interpreter, more explicit (*this is what we need to do*).

	ORG EN	SI EN DE
2:1	and Chinese people want what we have in the west ↲	und China euh strebt ja das westliche Entwicklungsmodell legitimerweise an ↲
	and they have every right to that ↲	
2:1	I was elected to this House twenty five years ago ↲	ich euh glaube wir ham heute / die wichtigste Debatte die hier stattfindet seit vierundzwanzig Jahren ↲
	this is probably the most important debate in which I've taken part ↲	

Table 6.10: Merged sentence alignment for ORG EN to SI DE EN.

The smallest category in the alignment analysis refers to cases where interpreters merge two or more source sentences into one sentence in the target (6.90% of alignment for English to German and 6.04% for German to English, cf. Tables 6.3 and 6.4). These merged sentences are mostly shorter at word level when compared to the combined source sentences they were based on (e.g. as in the examples in Table 6.10 with 17 words in the two source sentences combined rendered with only 9 words in one target sentence in the first example). However, they are usually as long as or longer than the individual source sentences. In addition to these differences in sentence length, it can also be stated that merged target sentences render the target output more dense. This can be seen in the first 2:1 example in Table 6.10, where the second source sentence *and they have every right to that* was translated by the adverb *legitimerweise* and simply included in the translation of the first source sentence, adding to the overall complexity of the target sentence. In the second example (Table 6.10), parts of the first source sentence were omitted or left implicit (the fact

that the speaker was elected), and only the time reference *twenty five years ago* is (wrongly) kept in the target sentence. These merged sentences can be expected to make the target text denser but only occur infrequently.

To summarise the findings of the alignment analysis, it can be stated that interpreters generally produce one target sentence for one source segment. However, the second most frequent category of the alignment analysis is 1:n alignments, where interpreters break up one source sentence into two or more target sentences. This leads to shorter and less complex target sentences, sometimes even using the same syntactic structure in several subsequent sentences. Furthermore, interpreters are found to omit source speech content within sentences (e.g. repetitions, adverbials, noun modifications such as adjectives) or omit typical sentences or sentence equivalents (e.g. Thank you) as well as occasionally omit complete content sentences (e.g. examples Table 6.9). It was found that syntactic restructuring at the macro-level is rare in interpreting. Interpreters only infrequently merge several source sentences into one target sentence, which can increase syntactic complexity in the target. Overall, the findings point toward more simplified output in interpreting than in original speech.

6.2.3 Dependency length analysis

In the following section, simplification is investigated via syntactic complexity measures, based on Dependency Locality Theory (DLT). Due to typological differences in the languages investigated (German and English), dependency length (DL) measures are not compared directly between the source and target languages. Instead, this section first only takes a comparable perspective and investigates English original speeches vs. interpreting into English from two different source languages (German and Spanish), and German original speeches vs. interpreting into German from English as source language. Nevertheless, general dependency measures on original speeches in the source language are also given when looking at target speeches and their comparable originals to investigate a potential source language influence (shining-through). For this, dependency measures are also given for the Spanish source speeches. In the following section, simplification is investigated via syntactic complexity measures, based on ependency locality theory (DLT). Due to typological differences in the languages investigated (German and English), dependency length (DL) measures are not compared directly between the source and target languages. Instead, this section first only takes a comparable perspective and investigates English original speeches vs. interpreting into English from two different source languages (German and Spanish), and German original speeches vs. interpreting into German from English as source language. Nevertheless, general dependency measures on original speeches in the source language are also given when looking at target speeches and their comparable originals to investigate a potential source language influence (shining-through). For this, dependency measures are also given for the Spanish

source speeches.

The quantitative analysis of syntactic complexity indicators uses corpus version EPIC UdS V3, (Stanza-parsed corpus version) as described in Section 3.1.4. First, the average dependency length (av DL), average maximum dependency length (av max DL) and adjacent dependencies (DL = 1) are compared for all tokens. In a next step, these sentences complexity indicators are investigated for short sentences and compared to long sentences. The focus of this analysis based on Dependency Locality Theory is an investigation of individual dependency relations and how they differ in interpreting and comparable originals with regard to their frequency and average length. The quantitative analysis uses a comparable approach, contrasting dependency relations in the originals and interpreting subcorpora and investigating which relations are more frequent and longer in interpreting vs. originals. Then these results are investigated in more detail in a parallel approach. This qualitative analysis will shed light on the potential reasons for these differences observed quantitatively.

Syntactic complexity indicators

The first general complexity indicator that is investigated for each subcorpus is average dependency length (av DL). Average dependency length is a measure that highly depends on the length of the individual sentences with longer sentences leaving more scope for longer dependency distances than shorter sentences. It is interesting to look at averages for each subset to assess overall complexity of each subset. Shorter average dependency length in a subset would make this subset less complex as a result. With interpreting using shorter sentences than comparable originals (cf. Section 6.2.1), interpreting should also have shorter average dependency length than originals. Table 6.11 gives an overview of average DL for each subcorpus.

subcorpus	N	mean	SD	median	IQR	range
ORG EN	3622	2.385497	0.6424033	2.357	0.7385	5.500
SI DE EN*	3661	*2.378656	0.6537530	2.312	0.7500	5.389
SI ES EN	3076	2.317782	0.6088192	2.286	0.7380	5.286
ORG DE*	3409	*2.792366	1.0272246	2.737	1.275	7.884
SI EN DE*	4076	*2.762564	0.9664613	2.714	1.165	7.682
ORG ES	2537	2.187714	0.6608731	2.192	0.738	5.711

Table 6.11: Average dependency length overview.

Generally, we can see shorter average dependency length (as indicated by *mean* and *median* in the table) for interpreting than for original speeches in the same language (comparable data). However, this trend is only statistically significant for English originals when compared to English interpreting based on Spanish source

speeches, and not when compared to English interpreted from German[67]. For the German dataset, the same trend can be observed with av DL and median DL shorter in interpreting than originals, but this is not statistically significant[8].

To put this result into perspective, DL of the source language, which may influence interpreters' production, must also be taken into account. For the significantly shorter interpreting output from Spanish as source language, it can be observed that original Spanish production has a shorter average DL than original English (cf.Table 6.11). The opposite is true for German originals with longer average DL than original English. Traces of these two source languages (longer av DL in original German, shorter av DL in Spanish) may be an explanation of only slightly shorter av DL in interpreting into English based on German, and significantly shorter av DL in interpreting with Spanish as source language.

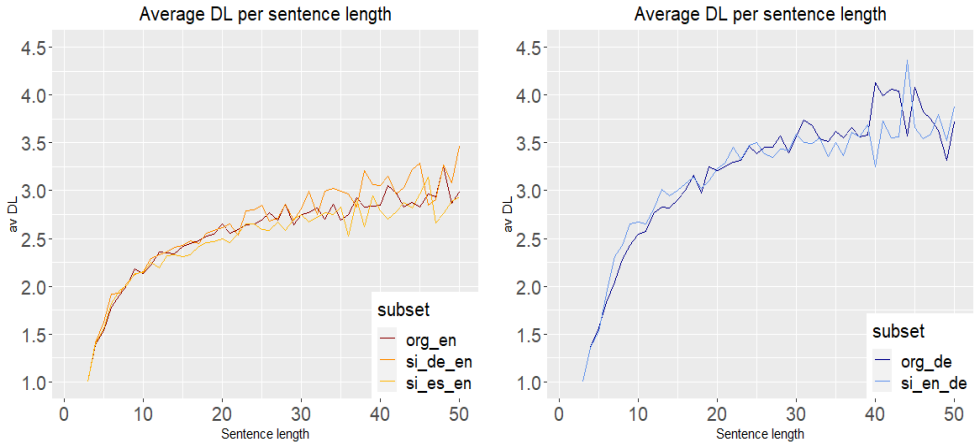


Figure 6.3: Average dependency length per sentence length for English (left) and German (right) comparable data.

As mentioned above, average sentence DL is a measure that highly depends on the overall length of a sentence, with longer sentences allowing for potentially longer dependency distances. Therefore, average DL is given for individual sentences length of 1 up to 50 tokens (Figure 6.3). For the comparable English dataset, it can be observed that short sentences of up to 15 tokens have no difference in average DL. Differences become visible for longer sentences, where we can see possible shining-through from longer average DL source sentences in original German and shorter average DL potentially influenced by shorter average DL in original Spanish.

DL of different languages cannot be compared directly. Nevertheless, Figure 6.4 plots comparable English data as well as the source speeches in one graph to visualize the influence of the source language shining trough. English interpreting from German

[6] * – not significant

[7] Welch Two Sample t-test for parametric data: ORG EN vs. SI DE EN: t = 0.45041, df = 7280.7, p-value = 0.6524; ORG EN vs. SI ES EN: t = 4.4225, df = 6616.1, p-value = 9.912e-06

[8] Welch Two Sample t-test for parametric data: ORG DE vs. SI EN DE: t = 1.284, df = 7077.7, p-value = 0.1992

(orange line) is closest to its German source (highest line in dark blue), while English interpreting from Spanish (yellow line) is the lowest line of the English dataset and closest to the Spanish original speeches (green line).

For the German dataset (Figure 6.3 on the right), differences in average DL can already be observed for short sentences, however, with interpreting showing higher average DL. A potential shining-through effect of the English sources cannot be the explanation (cf. Figure 6.5), and therefore, for short English source sentences interpreted into German, we cannot observe a trend for simplification. For longer sentences, German originals seem to have higher average DL than interpreting. However, as data get sparse for longer sentences, the trend is not very clear with both lines crossing several times. A more detailed analysis contrasting long vs short sentences will aim to give a clearer picture.

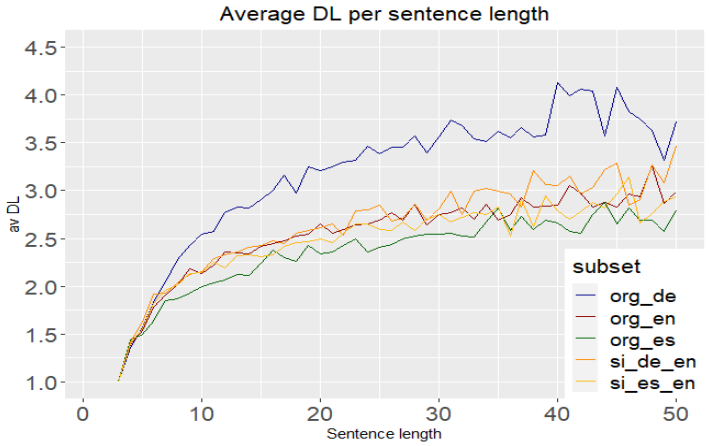


Figure 6.4: Average dependency length per sentence length for English comparable data and source speeches.

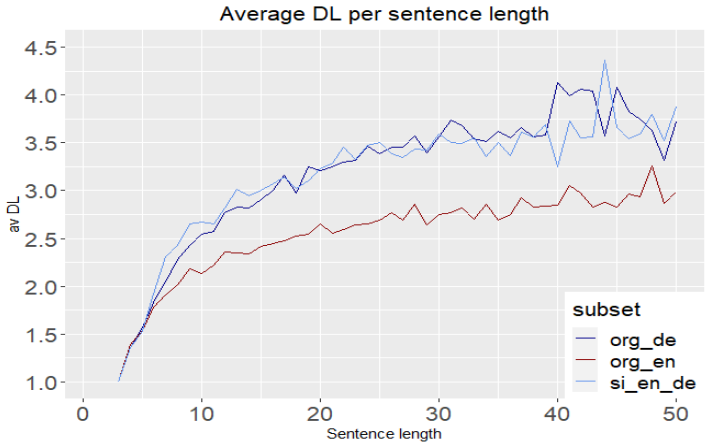


Figure 6.5: Average dependency length per sentence length for German comparable data and source speeches.

subcorpus	N	mean	SD	median	IQR	range
ORG EN	3622	7.783821	5.636698	6	6	44
SI DE EN	3661	7.179459	5.468366	6	5	67
SI ES EN	3076	7.136216	5.125988	6	5	59
ORG DE	3409	8.775887	5.712055	8	7	82
SI EN DE	4076	7.652846	4.540419	7	5	41
ORG ES	2537	8.317698	6.873517	7	7	65

Table 6.12: Maximum dependency length overview.

A further indicator of sentence complexity is average maximum DL (av max DL). Results for this indicator are given in Table 6.12 and Figure 6.6. We can see a clear trend of all interpreting subsets using lower average maximum DL than comparable originals in the same language, as displayed by lower mean values in Table 6.12 and additionally lower median values in interpreting for the German comparable dataset. This trend is statistically significant[9]. This result correlates with shorter sentences in interpreting than originals.

The boxplots in Figure 6.6 give a more detailed picture. The boxes for the interpreting values are generally lower and narrower than for comparable originals. Furthermore, the upper whiskers are also lower for interpreting than for originals. Therefore, we can see an overall trend of interpreting being more simple than comparable originals regarding this sentence complexity indicator. Furthermore, we can observe more outliers that are – at least for the English dataset – also higher in interpreting than in originals. This could point to traces of the generally higher (max) DL in the German and Spanish sources shining-through again.

subcorpus	N	av DL	av max DL	DL = 1
ORG EN	3622	2.385497	7.783821	38.7735 %
SI DE EN	3661	2.378656*	7.179459	38.1398 %
SI ES EN	3076	2.317782	7.136216	38.7493 %
ORG DE	3409	2.792366*	8.775887	40.3124 %
SI EN DE	4076	2.762564*	7.652846	39.6685 %
ORG ES	2537	2.187714	8.317698	44.9504 %

Table 6.13: Overview complexity indicators.

[9]Welch Two Sample t-test for parametric data: ORG EN vs. SI DE EN: t = 4.6434, df = 7268.8, p-value = 3.487e-06 ; ORG EN vs. SI ES EN: t = 4.9216, df = 6664.7, p-value = 8.789e-07; ORG DE vs. SI EN DE: t = 9.2852, df = 6454.2, p-value < 2.2e-16

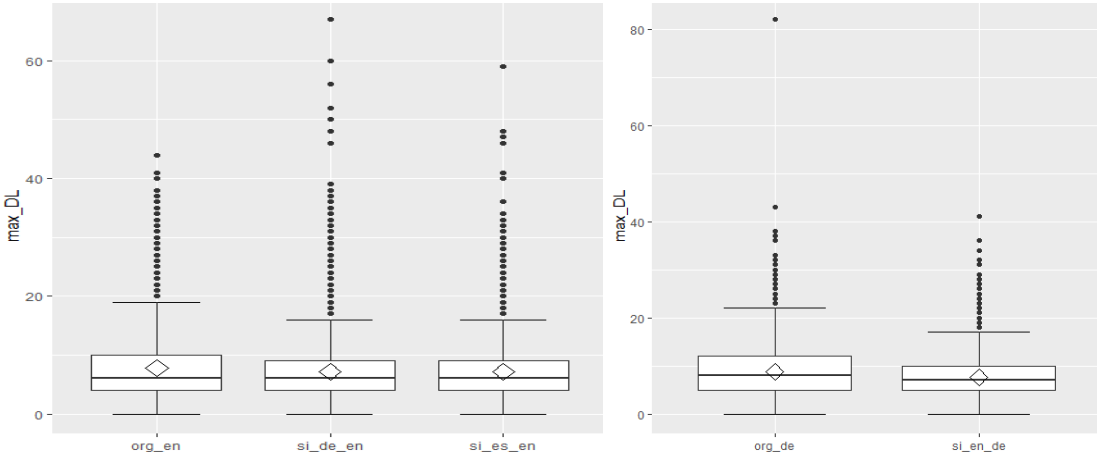


Figure 6.6: Maximum dependency length for English (left) and German (right) comparable data.

Table 6.13 summarises the sentence complexity indicators and adds the percentage of adjacent dependencies (DL = 1). Adjacent dependency relations are the shortest possible relations that do not have to be maintained in memory as the opened dependency relation is closed again immediately. A large share of these short relations makes it easier for a sentence to be processed. As interpreters generally produce fewer adjacent dependencies, it can be stated that, for this last complexity indicator, interpreting is less simple than comparable originals.

To summarise the sentence complexity indicators based on dependency length measures, it can be stated that the results are mixed. While maximum dependency length in interpreting is significantly lower than in comparable originals, the interpreting product cannot be seen as more simplified when looking at the number of adjacent dependencies. Generally, there seems to be a source language effect with (average) dependency distances being influenced by source language shining-through.

Short vs. long sentences

Short sentences have limited possibilities for restructuring of constituents. Therefore, minimal differences are expected for shorter sentences, while greater differences are expected in DL for longer sentences. Figure 6.3 shows no clear differences in the average dependency length for shorter sentences in the two languages studied, possibly even slightly shorter DL in German interpreting than German originals. Bigger differences appear for sentences longer than 35 tokens. However, data get sparser for longer sentences and therefore the plot line is less smooth. No clear pattern can be observed from the graph for longer sentences.

In the next step, short sentences are contrasted with long sentences. In order to picture an overall trend and handle the data sparsity problem for longer sentences, long sentences with length 41 to 45 are grouped together and contrasted with short

sentences (11 to 15 token). Table 6.14 shows the results for average dependency length, average maximum dependency length, and the percentage of adjacent dependencies (DL = 1) for short sentences with 11 to 15 tokens and long sentences with 41 to 45 tokens.

	short				long			
subcorpus	N	av DL	av max DL	DL = 1	N	av DL	av max DL	DL = 1
ORG EN	698	1.8911	5.44413	50.006%	98	2.5817	16.94898	49.110%
SI DE EN	875	1.8938	5.47086	49.947%	54	2.8338	19.44444	47.216%
SI ES EN	641	1.8420	5.27301	51.028%	65	2.7010	17.61539	48.294%
ORG DE	680	2.4996	7.78529	51.153%	55	3.7604	25.09091	48.176%
SI EN DE	969	2.5375	7.73787	50.620%	34	3.5756	23.55882	50.141%

Table 6.14: Dependency length indicators - short (11-41 tokens) vs. long (41-45 tokens) sentences.

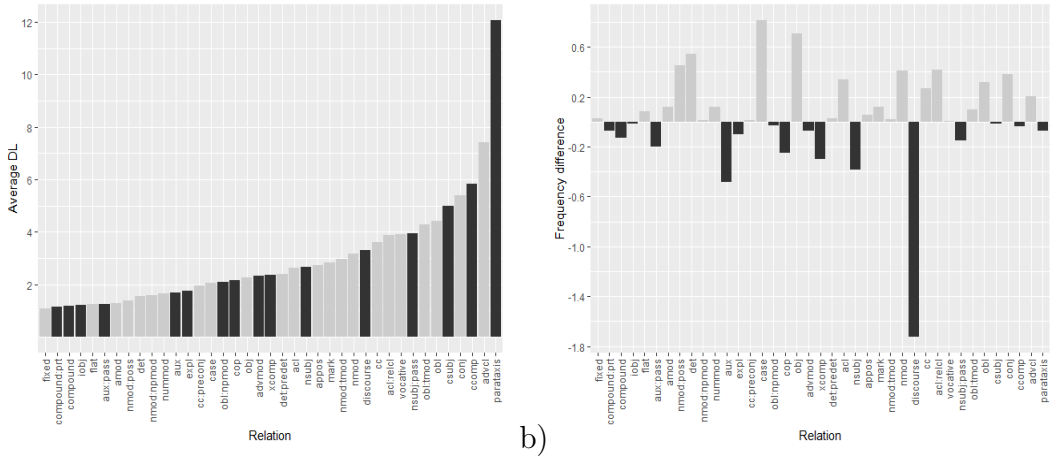
The results for German comparable data for average dependency length and average max dependency length confirm the hypothesis. Short sentences show similar results, whereas a trend of lower syntactic complexity with shorter dependency length (av DL, av max DL and DL=1) can be observed in long sentences. For English, greater differences can also be observed for the longer sentences. Contrary to German, English originals show lower syntactic complexity indicated by lower av DL, av max DL and DL = 1 than simultaneous interpreting into English from two source languages.

To conclude, a simplification trend can be observed for long sentences in German interpreting, but not for long sentences in English interpreting. This may be linked again to a source language influence in the interpreting product.

Frequency of dependency relations and average length

In order to find out which relations contribute most to the differences in syntactic complexity in interpreting vs. original speech, average DL as well as raw frequencies for each of the dependency relations are extracted from each subcorpus. To normalise the raw frequencies of relations, two approaches are taken: Figures 6.7, 6.8 and 6.19 take the total number of relations (equal to the number of tokens) in each subset as the basis for normalisation, while the number of sentences is used as the basis for normalisation in Figures 6.25, 6.26 and 6.27. While normalisation to the number of relations allows for comparing frequency differences overall, normalisation to the number of sentences in the subset considers that some relations only occur once in a sentence such as (such as the relation for predicate (root) or relations for the subject of the (main) clause, e.g. *nsubj, nsubj:pass, csubj* or *csubj:pass*). In both approaches, only relations with more than 5 occurrences are considered. In a next step, normalised frequencies for each dependency relation and subset are compared. If a relation occurs more frequently in the originals subset, this is depicted as a grey bar in the

charts. If a relation occurs more frequently in interpreting, the bar is visualised in black. The charts on the left (Figures a) visualise average dependency length of each relation, with the relations sorted according to increasing average dependency length. The charts on the right (Figures b) use the same order of relations; however, they visualise the frequency differences: Relations that are more frequent in interpreting are depicted in the negative range (again, visualised in black). Relations occurring more frequently in originals are visualised in the positive range of the y-axis, shown in grey. The length of the bars – regardless of depicted in the negative or positive range – shows the frequency differences[10]. In the following sections, interesting results of this comparison – particularly long relations and relations with large frequency differences – are investigated further.



a) b)

Figure 6.7: Dependency relations and their average dependency length (a) and their frequency differences (b) for ORG EN vs. SI DE EN. Grey denotes more frequent use in original, black marks more frequent use in simultaneous interpreting. Normalised to the number of relations.

For the English subsets (Figure 6.7 for original English compared to interpreted English with German as source language and Figure 6.8 for original English compared to interpreted English with Spanish as the source language), the longest relation – *parataxis* – is more frequent in interpreting when only considering Figures a) with the longest relation depicted in black. However, frequency differences for this relation are only minimal (only a short black bar for *parataxis* in Figures b). Furthermore, looking at the occurrences of *parataxis* in the English subsets shows that this relation is frequently annotated incorrectly: *parataxis* occurs frequently in the context of filled pauses (transcribed as *hum, hm* or *euh*) which are often linked to parsing errors such as in the example sentence, extracted from the corpus in Figure 6.9[11]. Figure 6.10 shows the correct parser output for the same sentence with filled pauses removed from

[10]The extracted frequencies and average DL used as basis for these visualisations is given in the Appendix (Tables A.10 - A.15).

[11]All figures use parser output of corpus version EPIC UdS V3. Parsing errors are not corrected unless specified otherwise.

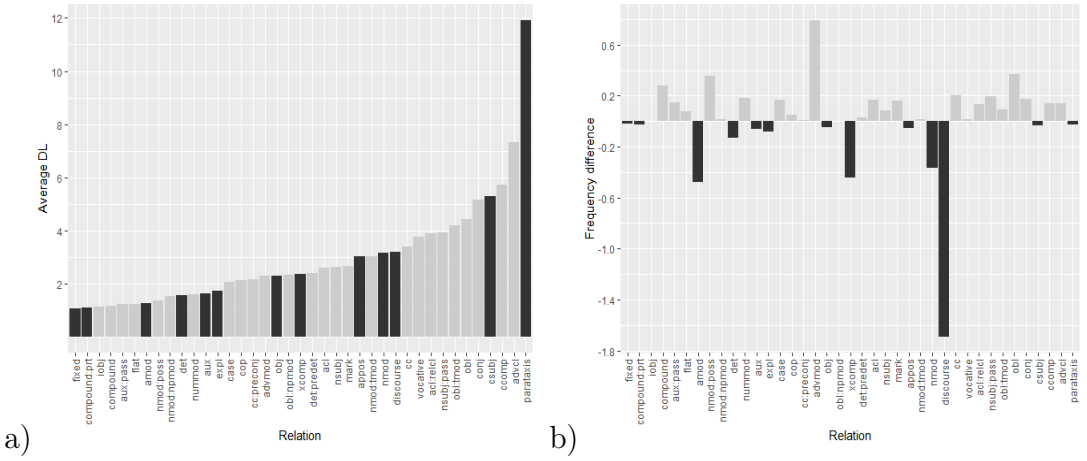


Figure 6.8: Dependency relations and their average dependency length (a) and their frequency differences (b) for ORG EN vs. SI ES EN. Grey denotes more frequent use in original, black marks more frequent use in simultaneous interpreting. Normalised to the number of relations.

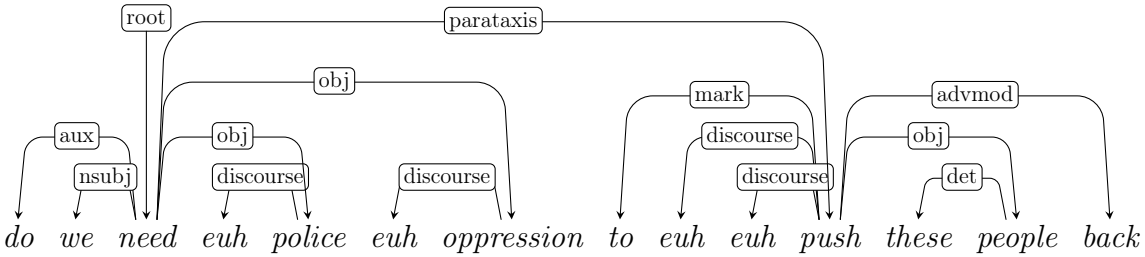


Figure 6.9: Example sentence with multiple filled pauses, leading to incorrect parsing output for the relation *parataxis* and *obj*.

the input. The correct parser output does not include the relation *parataxis* in this sentence. As filled pauses are a distinctive feature of interpreting (cf. Chapter 4), it can be expected that more filled pauses in interpreting lead to more incorrect parsing results in such contexts as shown with the incorrect use of the relation *parataxis* in interpreting in Figure 6.9. This could explain the observation in Figures 6.7 and 6.8 for this relation. As frequency differences are minimal and the related results potentially not very reliable, the relation *parataxis* is not considered further.

The largest frequency differences (Figures b) for any relation in both English subsets can be found for the relation *discourse* (discourse element), which occurs far more frequently in interpreting. This relation is generally used for POS INTJ, such as interjections, fillers, and non-adverbial discourse markers (de Marneffe et al., 2014). In the corpus variant EPIC UdS V3, filled pauses *hum, hm* and *euh*) were manually corrected to POS INTJ. An analysis of all dependency relations *discourse* in EPIC UdS EN shows that 95.05% of the relation *discourse* refer to filled pauses such as in the example in Figure 6.9. The frequency differences for this relation can therefore

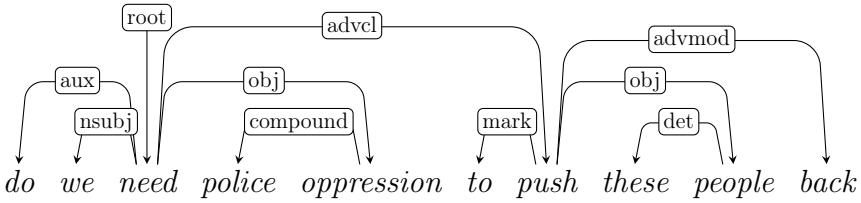


Figure 6.10: Correct parser output for example sentence without filled pauses.

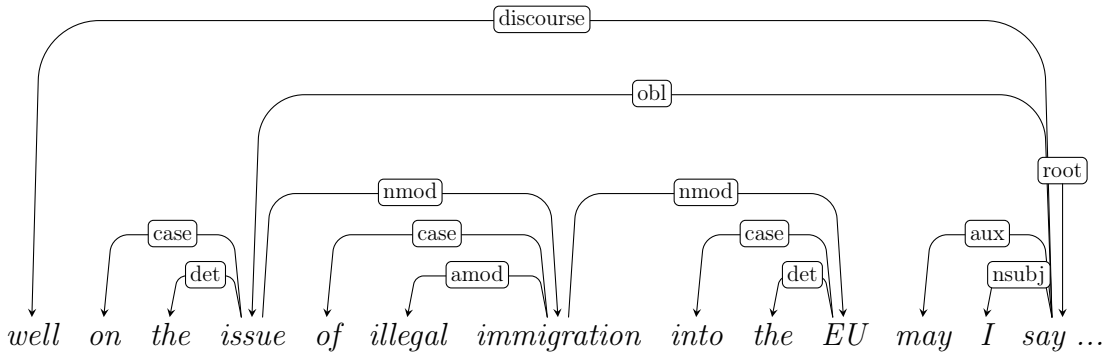


Figure 6.11: Example for relation *discourse*.

mainly be explained by the general observation that filled pauses are more frequently used in interpreting than originals (again, cf. Chapter 4). The dependency distance for filled pauses as *discourse* relation is generally very short (see the example in Figure 6.9). The remaining 4.05% of the relation *discourse* refer to discourse markers such as *well, yes* or *please* with generally longer dependency distances such as in the example in Figure 6.11. Such a use of the relation *discourse* is also in line with further observations made in Chapter 4, where the use of discourse markers was found to be distinctive for interpreting.

In addition to these two relations (*parataxis* and *discourse*) that are closely linked to the use of filled pauses, Figures 6.7 and 6.8 also point to more interesting observations. Especially longer relations (above av DL = 3) are used more frequently in original English than in both comparable English interpreting subsets. Frequency differences are particularly high for the long relations *advcl, conj* and *obl*, and therefore these relations are looked at in more detail.

The longest relation that is used more frequently in originals than in interpreting is the relation *advcl* – adverbial clause modifier (cf. Figure 6.7 and 6.8, Figures b, or Tables A.11 and A.13 in the Appendix). *advcl* refers to clausal dependents that modify a verb or other predicate, including temporal, consequence, conditional, or purpose clauses, etc. (de Marneffe et al., 2014). The example in Figure 6.12 shows the use of *advcl* in an adverbial clause of reason. Table 6.15 shows the full English original sentence and its interpretation into German. The German target is realised

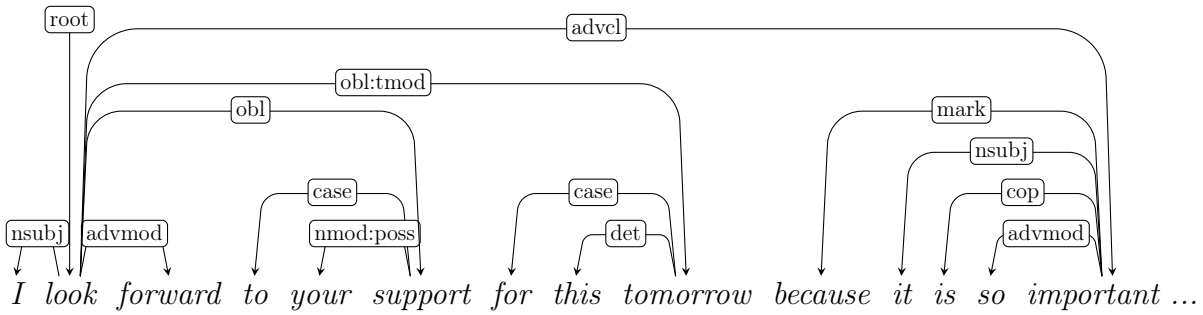


Figure 6.12: Example for relation *advcl*.

with two shorter independent clauses, not containing any further clausal relations.

ORG EN	SI EN DE
I look forward to your support for this tomorrow because it is so important *(advcl)* for all our futures ↲	ich freue mich auf die Abstimmung morgen ↲
	es ist sehr wichtig für unsere Zukunft ↲

Table 6.15: Example sentence for *advcl* in source (cf. Figure 6.12) and two independent clauses in the target.

The tendency of interpreters to split source sentences into several target sentences, as observed in Section 6.2.2, can be seen as a possible explanation for such clausal relations occurring more frequently in originals than in interpreting. Furthermore, clausal relations can easily be inferred and are frequently left implicit in translations (Hoek et al., 2017). Although the current analysis takes a comparable perspective and compares original English production with interpretations into English, the general tendency to split source sentences into several target sentences was observed for both interpreting directions and therefore can be used as a potential explanation for *advcl* being used more frequently in original English. Shorter and less syntactically complex sentences resulting from interpreters splitting sentences can generally be described as simplification in *interpretese*.

The second long relation that occurs more frequently in English originals than in interpreting is the relation *conj* (conjunct), which refers to a relation between two elements that are connected by a coordinating conjunction. The head of the relation is the first conjunct. All other conjuncts depend on the head via the *conj* relation (de Marneffe et al., 2014). This refers to coordination below the clause level (such as coordination of noun phrases, adjective phrases, etc.) and the coordination of clauses. As stated in the segmentation guidelines (Section 3.1.2), multiple coordinated main clauses, apart from elliptical ones, that are linked by a coordinating conjunction, are treated as separate segments or sentences. Therefore, the dependency relation *conj* in EPIC UdS V3 does not relate to coordination of independent clauses, but

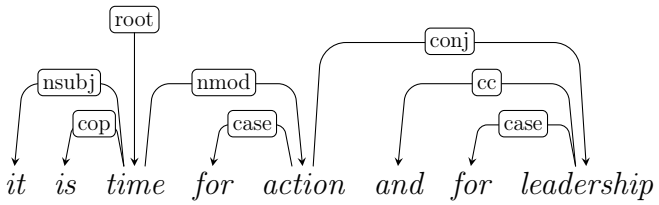


Figure 6.13: Example for relation *conj*, coordination of noun.

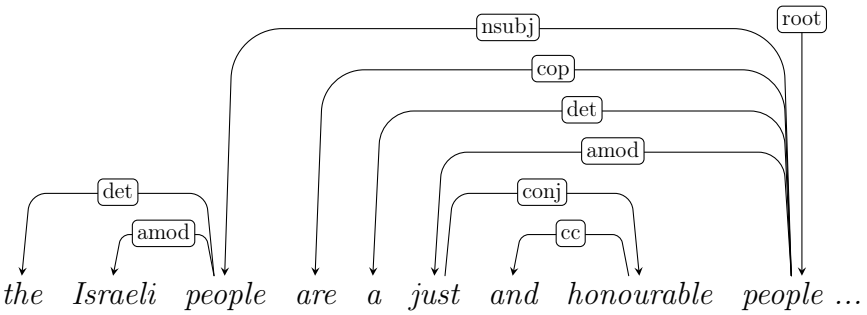


Figure 6.14: Example for relation *conj*, coordination of adjective.

rather to coordination below the clause level, such as coordinated nouns (example in Figure 6.13), adjective phrases (example in Figure 6.14), etc. or coordination of clauses with the same subject (example in Figure 6.15).

The example sentence in Figure 6.13 from the original English subset shows coordination of two nouns, *action* and *leadership*. The interpretation into German (Table 6.16) gives a combined and verbal rendition of the source, which does not include coordination.

ORG EN	SI EN DE
it is time for action and for leadership (*conj*) ↲	wir brauchen wirklich euh ein/ euh einer der die Sachen in die Hand nimmt ↲

Table 6.16: Example sentence for *conj* in source (cf. Figure 6.13), coordination of noun and no coordination in the target.

The coordination in the example in Figure 6.14 is part of the main clause in the English original sentence. The interpreted version (Table 6.17) shows, like in the previous example, an omission of source speech input, although in this case the entire adjective group is omitted and the interpreted version structurally less complex.

Both examples are in line with the results presented in Przybyl et al. (2022a), where the authors showed that interpreters tend to produce shorter noun phrases, potentially leaving out (optional or redundant) noun modifiers, and therefore producing a simpler target text.

The relation *conj* below clause level generally refers to relatively short dependency

ORG EN	SI EN DE
the Israeli people are a just and honourable (*conj*) people who have suffered miserably throughout the centuries in this continent ↲	die Israelis sind ein Volk das über Jahrhunderte lang euh gelitten haben ↲

Table 6.17: Example sentence for *conj* in source (cf. Figure 6.14), coordination of adjective and no coordination in the target.

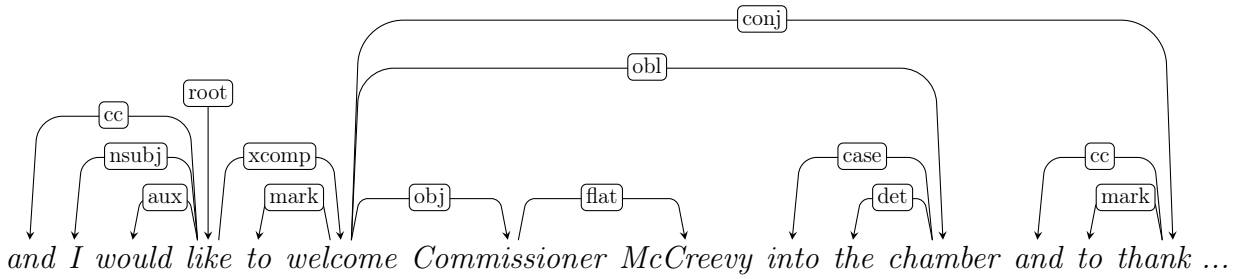


Figure 6.15: Example for relation *conj*, clausal use in original English.

relations. In the example in Figure 6.14, *honorable* and its head *just* are two words apart (DL = 2), the distance for *conj* in the example in Figure 6.13 is DL = 3 (distance between *action* and *leadership*). Memory effort to retain these items in memory until the dependency relation is closed is fairly low. If these items are left out in the target production, it may be more due to general cognitive load and time pressure in interpreting, rather than effort to maintain these relations in memory.

Longer *conj* relations refer mainly to coordinated clauses (Figure 6.15). As stated previously, interpreters tend to split such coordination into several independent clauses as in Table 6.18.

ORG EN	SI EN DE
and I'd like to welcome Commissioner McCreevy into the chamber and to thank (*conj*) all colleagues who are here on this evening ↲	ich bin auch sehr froh dass Kommissar Herr Creevy zu uns gekommen ist ↲
	ich möchte auch alle anderen Kollegen begrüßen ↲

Table 6.18: Example sentence for *conj* in source (cf. Figure 6.15), coordination of clause and no coordination in the target.

The example in Figure 6.16 contains three clausal *conj* relations, which are reduced to only two clausal relations in the German interpreted rendition (*xcomp* and *conj*) by omitting the last clause *take them home to eat* completely (Figure 6.17 and Table 6.19). Again, this analysis focusses on a comparable perspective. However,

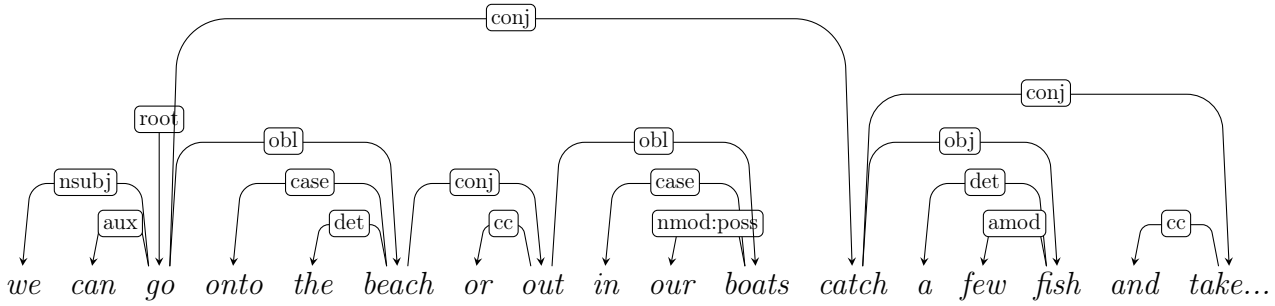


Figure 6.16: Example for relation *conj*, clausal use in original English.

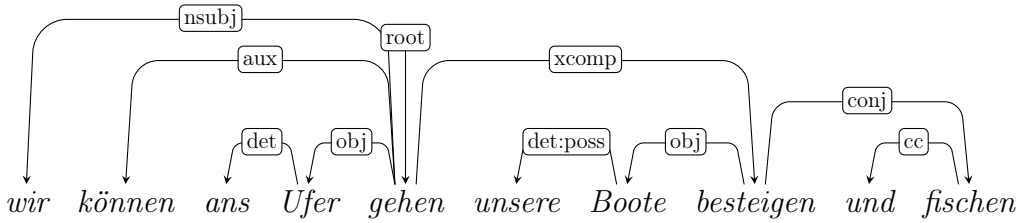


Figure 6.17: Interpreted sentence for relation *conj*.

omission in interpreting, as shown by comparing source and target, may be another reason for some relations to be less frequent in interpreting.

ORG EN	SI EN DE
we can go onto the beach or out (*conj*) in our boats catch (*conj*) a few fish and take (*conj*) them home to eat ↲	wir können ans Ufer gehen unsere Boote besteigen (*xcomp*) und fischen (*conj*) ↲

Table 6.19: Example sentence for multiple *conj* in source (cf. Figure 6.16) and only one coordination in the target.

The last long relation looked at in more detail, also occurring more frequently in original English than comparable English interpreting, is *obl* (oblique nominal). This relation is used for a noun, pronoun, or noun phrase acting as an oblique (non-core) argument or adjunct. Functionally, it corresponds to an adverbial attaching to a verb, adjective, or other adverb (de Marneffe et al., 2014).

The example in Figure 6.18 shows the use of the relation *obl* connecting an adverbial of time and an adverbial of place to the verb of the relative clause. The entire English original sentence and the interpreted version are given in Table 6.20. Besides the omission of *just and honorable* as discussed above, the German interpreted version contains further omissions: the German sentence only uses the adverbial of time (*throughout the centuries - über Jahrhunderte lang*) and omits the adverbial of place (*in this continent).* As the relation *obl* refers to relations that are optional in a clause, interpreters, who generally produce shorter sentences, may occasionally

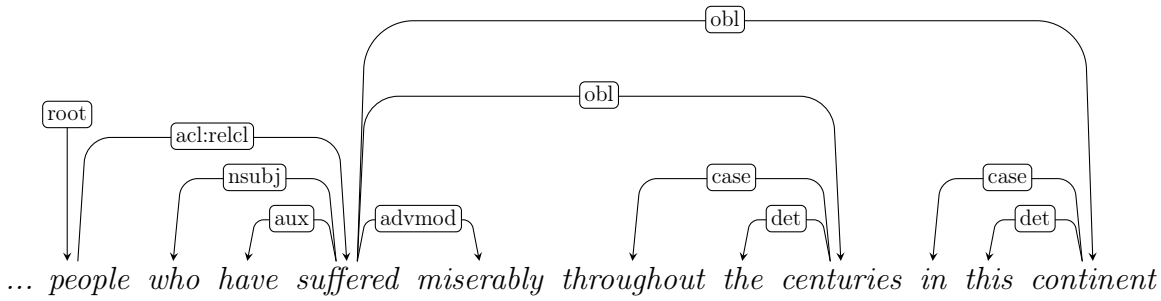


Figure 6.18: Example for relation *obl*.

ORG EN	SI EN DE
the Israeli people are a just and honourable people who have suffered miserably throughout the centuries (*obl*) in this continent (*obl*) ↲	die Israelis sind ein Volk das über Jahrhunderte (*obl*) lang euh gelitten haben ↲

Table 6.20: Example sentence for *obl* in source (cf. Figure 6.18) and multiple omissions in the target.

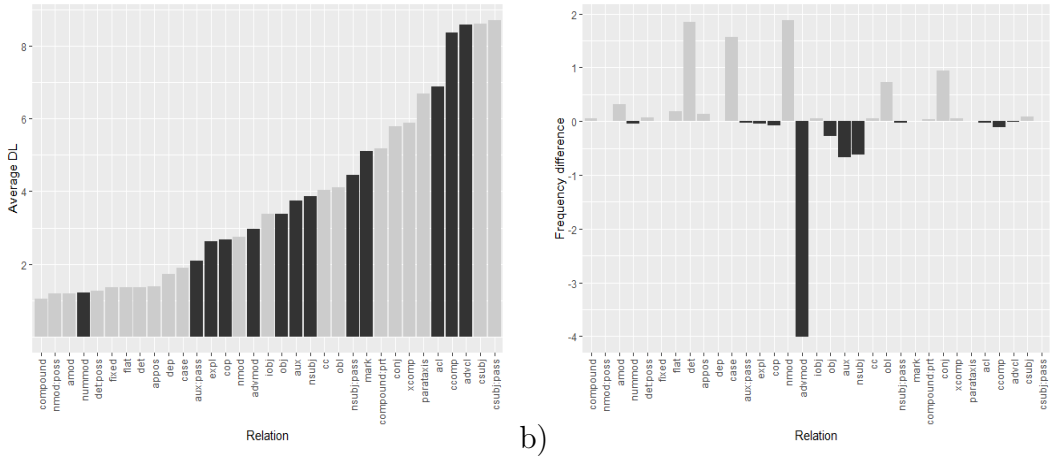
omit such (syntactically) optional elements – although present in the source – in their target rendition.

The observations made so far for the relations discussed above are observations made for both English comparable datasets. English originals compared to English interpreting from German as source and English originals compared to English interpreting from Spanish as source. However, it should be noted that the source language also has an effect on frequencies and frequency differences of dependency relations. As can be observed in Figures 6.7 and 6.8, the long relation *nsubj:pass* is found more frequently in interpreting from German compared to original English, while it is more frequent in English originals when compared to interpreting from Spanish.

To summarise the findings for average length and frequency of dependency relation in the English comparable dataset, it can be stated that long clausal relations tend to be more frequent in original English than in interpreted English. Long relations are potentially more challenging to produce, as open relations have to be retained in memory until they are resolved. It is likely that interpreters try to avoid long relations and therefore tend to split source sentences into several independent sentences as well as omit source content. However, the source language also has an effect on the use and length of dependency relations in interpreting.

Looking at frequency differences and average length of dependency relations in the German comparable datasets shows a slightly different picture (Figure 6.19). The observation made for the English dataset that long dependency relations occur more frequently in originals cannot be confirmed. In the German dataset, frequency

differences, especially for long relations, are minimal. Nevertheless, several relations also show large frequency differences. One relation is used much more frequently in German interpreting (*advmod*), whereas several relations are more frequent in German originals (*det, case, nmod, obl*, and *conj*).



a) b)

Figure 6.19: Dependency relations and their average dependency length (a) and their frequency differences (b) for ORG DE vs. SI EN DE. Grey denotes more frequent use in original, black marks more frequent use in simultaneous interpreting. Normalised to the number of relations.

The relation *advmod* is used for an adverbial modifier of a word. It refers to a non-clausal adverb or adverbial phrase that modifies a predicate or a modifier word (de Marneffe et al., 2014). In the German dataset, it relates to a medium-length relation that is used much more frequently in interpreting (15.73% of relations) than in originals (11.71% of relations).

These frequency differences may be partly explained by more filled pauses in interpreting overall, which are tagged *advmod*. 15.66% of all *advmod* in the German dataset relate to filled pauses *euh, hum* and *hm*. An example of such a use of *advmod* is given in Figure 6.20. A similar observation was made for the English dataset with filled pauses assigned to the relation *discourse* with more frequent use in English interpreting. However, filled pauses account for only 15.66% of all relations *advmod* in the German dataset. Moreover, the relation *advmod* also relates to other modifiers

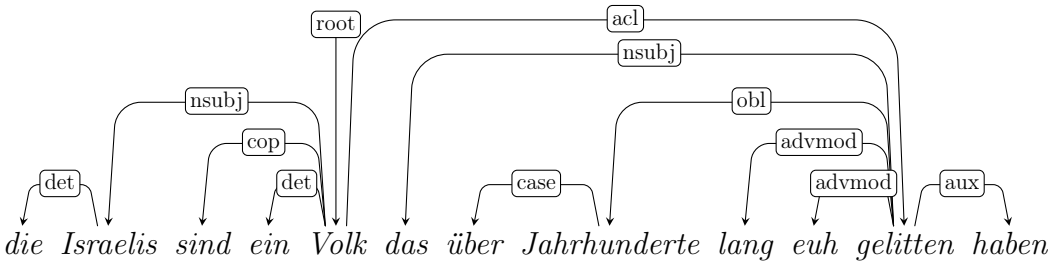


Figure 6.20: Example for relation *advmod*.

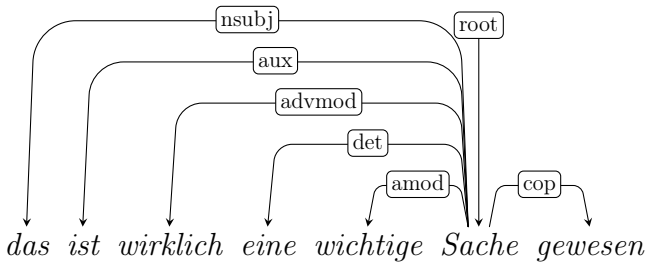


Figure 6.21: Example for relation *advmod* - intensifier.

such as *auch, wirklich, so, hier, jetzt*, or *ja*. Such modifiers (used as intensifiers, discourse markers, and deictics) were also found distinctive for interpreting by means of relative entropy in Section 4.2. The result of *advmod* occurring more frequently in interpreting than in comparable German originals thus confirms the observations made in Section 4.2. The example sentence in Figure 6.21 shows the *advmod* relation in the use of an intensifier (*wirklich*), which was added in interpreting and was not used in the English source input. The English original sentence this interpreted sentence was based on (cf. Table 6.21) does not include such an intensifying adverbial modification of the noun. The same tendency can be observed in the example in Table 6.22, with two relations *advmod (unbedingt, einmal)*, which are not triggered by source speech content. The source message is fairly weak (I think), whereas *unbedingt* (EN: absolutely) in the target makes the instructions given in the target more obligatory. This is only slightly weakened by the addition of the second *advmod* (*einmal*, EN: some day). More frequent use of the relation *advmod* in interpreting may therefore be partly due to more frequent use of filled pauses in interpreting, but may also be linked to additions (intensifiers, deictics) by the interpreter.

ORG EN	SI EN DE
and it's a great exercise ↲	das ist wirklich (*advmod*) eine gute Sache gewesen ↲

Table 6.21: Example sentence for *advmod*, intensifier use in target (cf. Figure 6.21) and no intensifier in the source.

Such a use of this dependency relation generally relates to fairly short dependency distances, which do not have to be maintained in memory for long, if at all (e.g. discourse markers at the beginning of the sentence, intensifier directly before noun). Therefore, it can be assumed that the more frequent use of *advmod* generally does not make the interpreting product more difficult to be processed.

A second relation that occurs more frequently in German interpreting than in originals is *aux*: an auxiliary of a clause is a function word associated with a verbal predicate that expresses tense, mood, aspect, voice, or evidentiality (de Marneffe et al., 2014). Although frequency differences are not as large as for the relation

ORG EN	SI EN DE
and I think we have to look at the euh the career structures that we provide ↲	wir müssen uns unbedingt (*advmod*) einmal (*advmod*) die Karriere und Laufbahnmöglichkeiten anschauen ↲

Table 6.22: Example sentence for *advmod*, intensifier use in target and no intensifier in the source.

advmod, differences are still noticeable with *aux* amounting to 4.14% of the relations in interpreting and only 3.47% in originals.

One use of auxiliaries in German can be linked to the past tense. The German tense *perfect (Perfekt)* is formed using an auxiliary verb plus a past participle. Alternatively, the *simple past (Präteritum*) may be used in many cases. However, there are stylistic differences. The past tense using an auxiliary and past participle can be linked to spoken language, informal use, or colloquial language, while *Präteritum* is used more in a written or more formal context (Schneider and Lang, 2022). The example in Table 6.23 shows such spoken use of *Perfekt* that alternatively could have been phrased as *...das über Jahrhunderte lang litt* in *Präteritum* tense. Looking at the source sentence on which it was based, shows that in addition to the spoken influence, the tense used in the English original sentence may also be a reason for choosing this verb form: *have + past participle* (present perfect) is translated into *haben + past participle*. Choosing a different past tense in German may have required more cognitive resources than simply choosing a translation that is structurally very close to the original it was based on (cf. the use of cognates in Section 5.2.4).

ORG EN	SI EN DE
the Israeli people are a just and honourable people who've (*aux*) suffered miserably throughout the centuries in this continent ↲	die Israelis sind ein Volk das über Jahrhunderte lang euh gelitten haben (*aux*) ↲

Table 6.23: Example sentence for *aux*, triggered by source.

ORG EN	SI EN DE
I just highlight two points Madam President ↲	zwei Dinge möchte (*aux*) ich hervorheben Frau Präsidentin ↲

Table 6.24: Example sentence for *aux*, not triggered by source.

However, not all occurrences of auxiliaries in interpreting are triggered by a source auxiliary. The example in Table 6.24 shows the addition of a modal auxiliary to the target speech. These observations are again in line with the results analysed in Section 4.2 with modals and auxiliaries being more distinctive for interpreting than for

originals and in line with Shlesinger and Ordan (2012) findings of interpreting using more spoken features. More frequent use of *aux* may be a combination of source language shining-through (such as from present perfect use in the English original speeches on which the interpreted speeches were based), but also (over)normalisation – additions that make the target text more polite.

One further relation occurs more frequently in German interpreting than in originals: *nsubj* (nominal subject) refers to the syntactic subject and proto-agent of a clause. This nominal may be headed by a noun, pronoun, or relative pronoun. In ellipsis contexts, the head may also be an adjective(de Marneffe et al., 2014).

ORG EN	SI EN DE
and here we (*nsubj*) are again in the MFF looking at humanitarian aid budget and the development budget and prevaricating whether we (*nsubj*) 're going to meet the commitment we (*nsubj*) 've already entered into ↲	da müssen wir (*nsubj*) auch die Engagements einhalten die wir (*nsubj*) eingegangen sind ↲

Table 6.25: Example sentence for *nsubj*.

The example in Table 6.25 shows a long source sentence (32 words) including 3 relations *nsubj*. The interpreted sentence is a condensed summarising rendition of the source input, including 2 *nsubj* but only 11 words. When normalising to the number of tokens, this shortening of source input therefore leads to a higher frequency of *nsubj* overall in interpreting (32/3 vs. 11/2).

Looking at dependency relations that are used more frequently in German originals than in German interpreting, the following observations can be made. Two of these relations are long relations that are also more frequently used in English originals than English interpreting: *conj* and *obl*.

The relation *conj* (conjunct), as stated above for the analysis of the English dataset, relates to two elements that are connected by a coordinating conjunction (e.g. und, oder, sondern or bzw). The first conjunct is treated as the head of the relation. All other conjuncts depend on the head, labelled as *conj* (de Marneffe et al., 2014).

As for the English dataset investigated above, shorter relations *conj* below clause level are used for coordination of phrases such as in Table 6.26. In this example, *conj* in the source relates to coordination within the subject noun phrase. Such use of *conj* is often also reproduced in interpreting. These shorter relations can be easily retained in memory for a short period of time.

Coordination of clauses relates to longer dependency distances that potentially require more cognitive resources to be retained in memory. Such uses of *conj* that

ORG DE	SI DE EN
auch Pressefreiheit und Minderheitenschutz (*conj*) aber auch die Reform (*conj*) der Justiz müssen permanent überprüft werden ↲	euh pro/ protection of minorities freedom (*conj*) of the press and reform (*conj*) of the judiciary this is something that needs to be monitored in a ongoing way ↲

Table 6.26: Example sentence for *conj*, coordination of phrases in source and target.

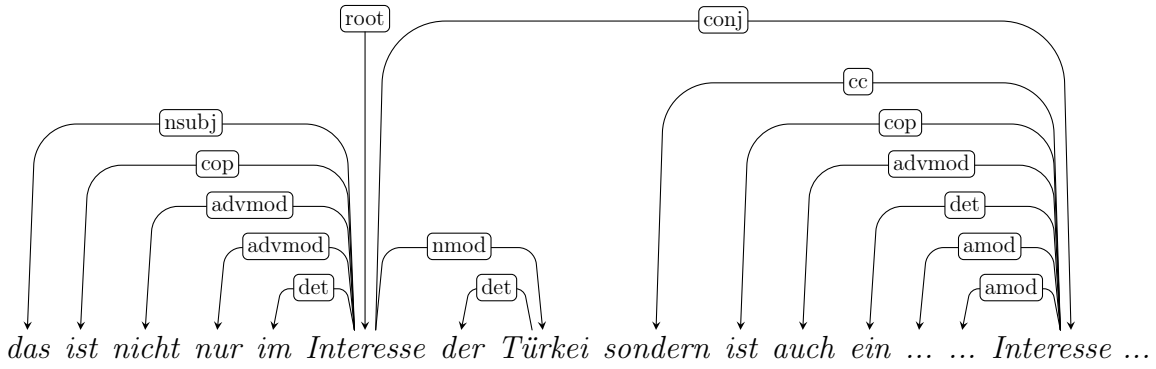


Figure 6.22: Example for relation *conj* - coordination of clauses.

are realised in one sentence in the source (Figure 6.22) tend to be split into several independent clauses in the target (Table 6.27).

ORG DE	SI DE EN
das ist nicht nur im Interesse der Türkei sondern ist auch ein wichtiges strategisches Interesse (*conj*) der Europäischen Union ↲	that's not only something which is of interest to Turkey ↲
	it's also / important ↲
	and it is in the interest of the European Union ↲

Table 6.27: Example sentence for *conj*, coordination of clauses in source (cf. Figure 6.22) and independent clauses in target.

In line with results for the English dataset, the second relation that is more frequent in German originals than in interpreting into German is the non-core dependent *obl* (oblique). It is used for nominal adjuncts or non-core arguments, such as non-core noun phrases in dative or genitive, or nominal adjuncts in prepositional phrases (de Marneffe et al., 2014).

The example given in Figure 6.23 and Table 6.28 shows such an optional argument in the source *in Summe*, which is omitted in the interpreted rendition.

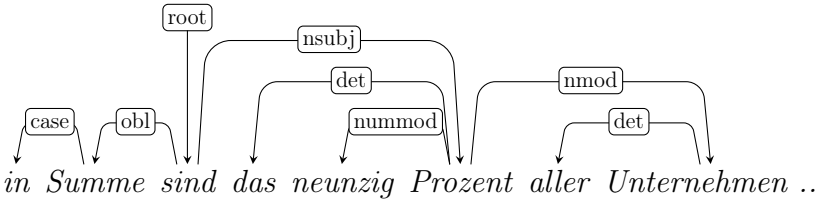


Figure 6.23: Example for relation *obl*.

ORG DE	SI DE EN
in Summe (*obl)* sind das neunzig Prozent aller Unternehmen die unter zehn Arbeitnehmern in Österreich beschäftigen ↲	euh and ninety per cent of all companies in Austria hire less than ten people to give you an idea ↲

Table 6.28: Example sentence for *obl* cf. Figure 6.23), oblique argument omitted in target.

Further relations that are used more frequently in German originals than in interpreting into German are relations that can be linked to the observation made in Section 4.2 of originals being more nominal than comparable interpreting: *det, case, nmod*.

The relation between a noun and its determiner is labelled *det* (determiner). This includes indefinite, demonstrative, or interrogative pronouns. *case* (case marking) is mostly used for prepositional phrases where prepositions are regarded as dependents of the noun whose case value they govern. *nmod* (nominal modifier) refers to the nominal dependents of nouns or noun phrases (de Marneffe et al., 2014).

ORG DE	SI DE EN
und die Wahlfreiheit der (*det*) Mitgliedstaaten (*nmod*) in (*case*) diesen (*det*) Angelegenheiten (*nmod*) schwächt den Binnenmarkt ↲	and the freedom of (*case*) Member States (*nmod*) would weaken the internal market ↲

Table 6.29: Example sentence for *det, case* and *nmod* (cf. Figure 6.24), partially omitted in target.

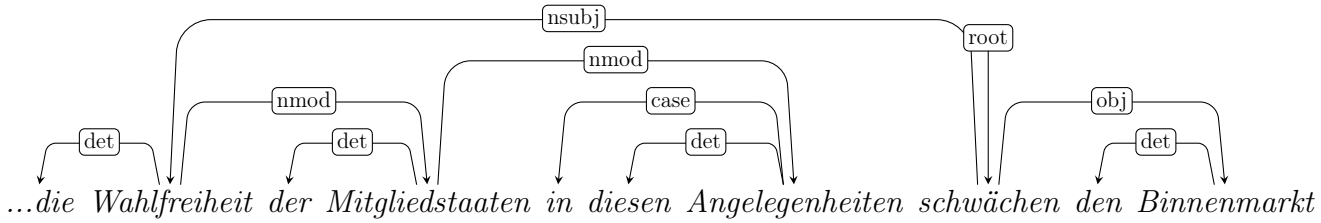
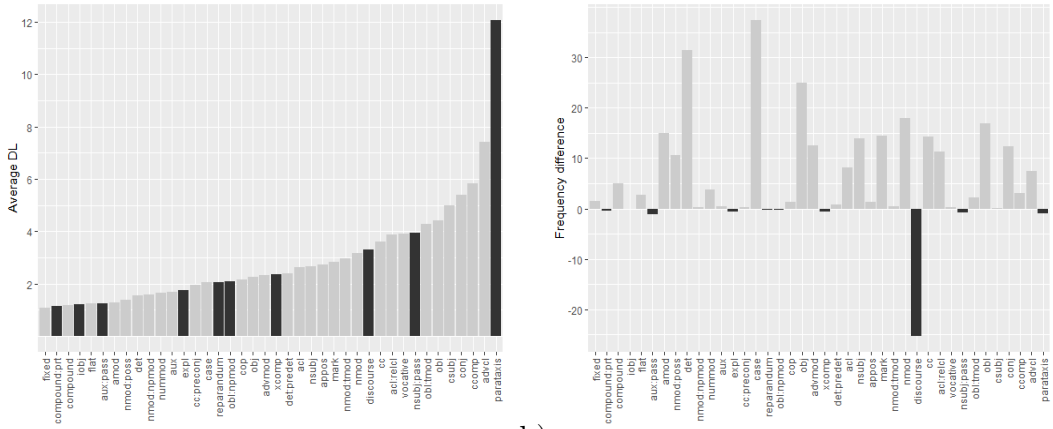


Figure 6.24: Example for relations *det, case* and *nmod*.

The example in Figure 6.24 shows these relations combined in the nominal subject of the sentence. It is an example taken from the German originals dataset. The subject head (*Wahlfreiheit*) is nominally post-modified with *der Mitgliedstaten* and *in diesen Angelegenheiten* in the source. The second post-modification is omitted in the interpreted version (cf. Table 6.29). By omitting one of the nominal modifications, the target becomes less dense at the lexical level and less complex syntactically.

As a summary for the German comparable dataset regarding average length and frequency of dependency relations, it can be stated that the analysis mainly confirms the observations made in Sections 4.2: German originals are characterised by nominal relations, whereas German interpreting uses spoken features more frequently (filled pauses, intensifiers, and auxiliaries). Interpreters' tendency to split source sentences (cf. Section 6.2.2) can be confirmed by more clausal relations in originals.

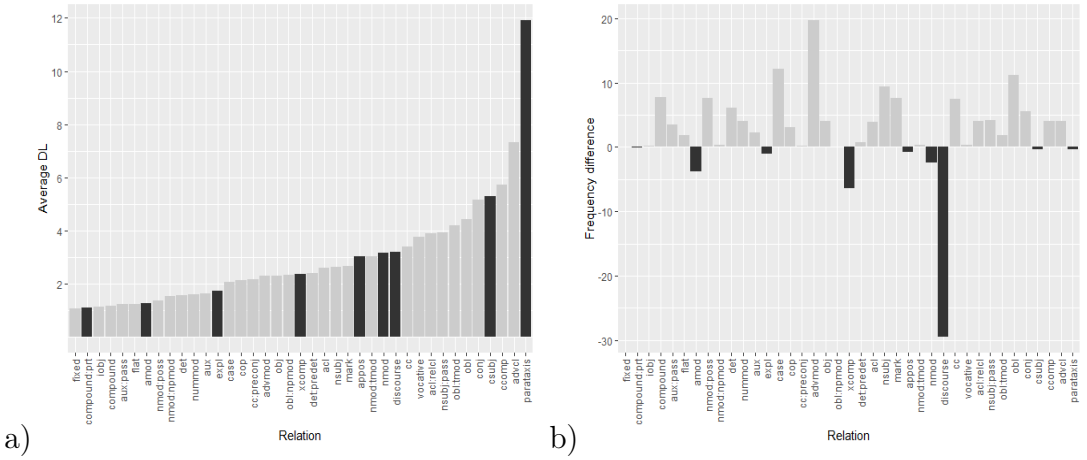
While the approach used so far looked at frequency differences and average length of dependency relations normalised to the total number of relations, Figures 6.25, 6.26 and 6.27 use frequency differences for each relation normalised to the number of sentences. When analysing these results, it has to be considered that sentences in originals tend to be longer than in interpreting (cf. Section 6.2.1) and more relations occur in longer sentences. However, as the relations used may differ (core vs. non-core relations), it is still interesting to investigate the results on sentence level further.



a) b)

Figure 6.25: Dependency relations and their average dependency length (a) and their frequency differences (b) for ORG EN vs. SI DE EN. Grey denotes more frequent use in original, black marks more frequent use in simultaneous interpreting. Normalised to the number of sentences.

For the English dataset normalised to the number of sentences (Figure 6.25 for ORG EN vs. SI DE EN and Figure 6.26 for ORG EN vs. SI ES EN), previously made observation of longer relations occurring more frequently in originals than in interpreting can be confirmed for both comparable English datasets. The effect is

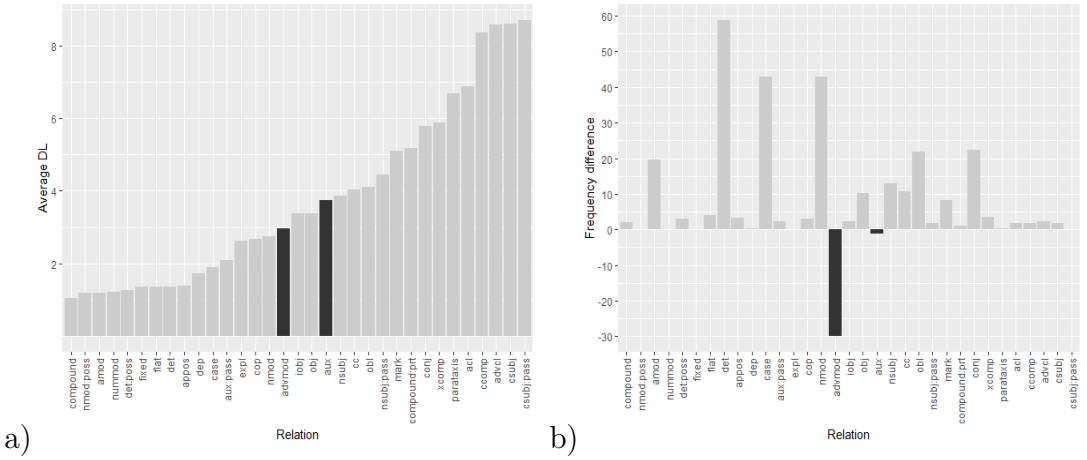


a) b)

Figure 6.26: Dependency relations and their average dependency length (a) and their frequency differences (b) for ORG EN vs. SI ES EN. Grey denotes more frequent use in original, black marks more frequent use in simultaneous interpreting. Normalised to the number of sentences.

even stronger than for normalisation to the number of relations (higher light grey bars depicting frequency differences in the chart on the right - Figures b)). An exception, as discussed above, are *parataxis* and *csubj* for SI ES EN, with only minor frequency differences.

Due to longer sentences in originals, most relations occur more frequently in originals. A clear exception, also discussed above, is the relation *discourse*, linked to filled pauses that occur more frequently in interpreting.



a) b)

Figure 6.27: Dependency relations and their average dependency length (a) and their frequency differences (b) for ORG DE vs. SI EN DE. Grey denotes more frequent use in original, black marks more frequent use in simultaneous interpreting. Normalised to the number of sentences.

For the German comparable dataset (Figure 6.27) with relations normalised to the number of sentences, it can be stated that all but two relations occur more frequently

in originals (*advmod* and *aux*). Both relations, as discussed above, can be linked to interpreting using more spoken elements than original spoken.

To summarise the analysis of individual dependency relations, their length, and frequencies across the subsets, it can be stated that clausal relations tend to be more frequent in originals of both languages studied. Clausal relations tend to be long, and interpreters seem to try to avoid the additional load on memory which needs to be invested to maintain open dependency relations in memory by splitting source sentences into shorter target sentences or omitting content from the source.

Nominal relations tend to be more frequent in originals, whereas relations that can be linked to spoken features are used more frequently in interpreting. Both observations are in line with previous findings within this thesis that originals are more nominal and interpreting uses more spoken elements (Section 4.2) and (Shlesinger and Ordan, 2012). Furthermore, the observations are also in line with the results from Section 6.2.2, with the splitting of the source input leading to shorter sentences and shorter dependency relations. Additionally, a trend of summarising and omissions was observed.

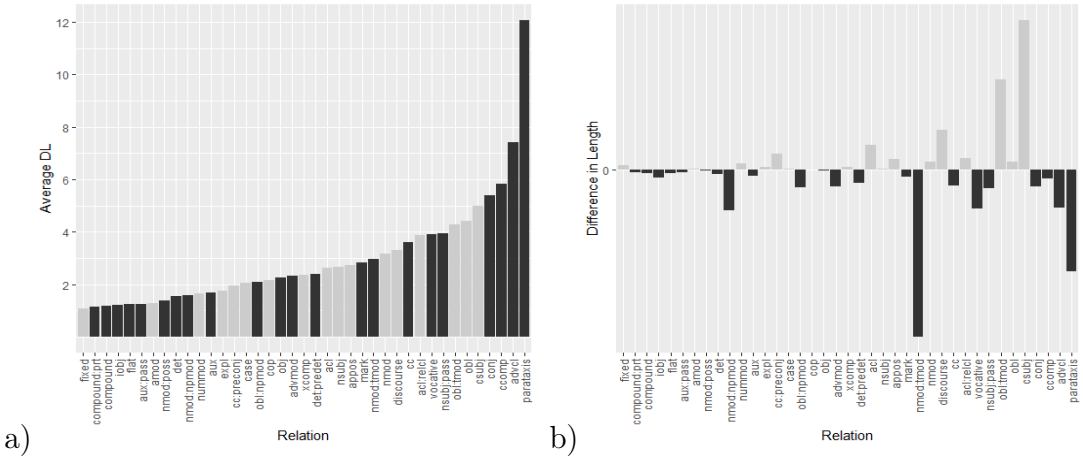
Analysing interpreters' output from two different source languages revealed some differences in the use of dependency relations; however, the overall trends as described above are confirmed.

Average length of dependency relations

While the previous section focused on frequency differences for individual dependency relations and their average length, this section investigates the average length of relations and the question of whether individual relations are longer on average in original speeches or interpreting of the same language. This is particularly relevant because it investigates the question whether interpreters tend to produce shorter dependency relations to reduce the load on cognitive resources. While the previous investigation assesses which relations are used more frequently, this analysis tries to investigate whether interpreters produce shorter distances for the same relations.

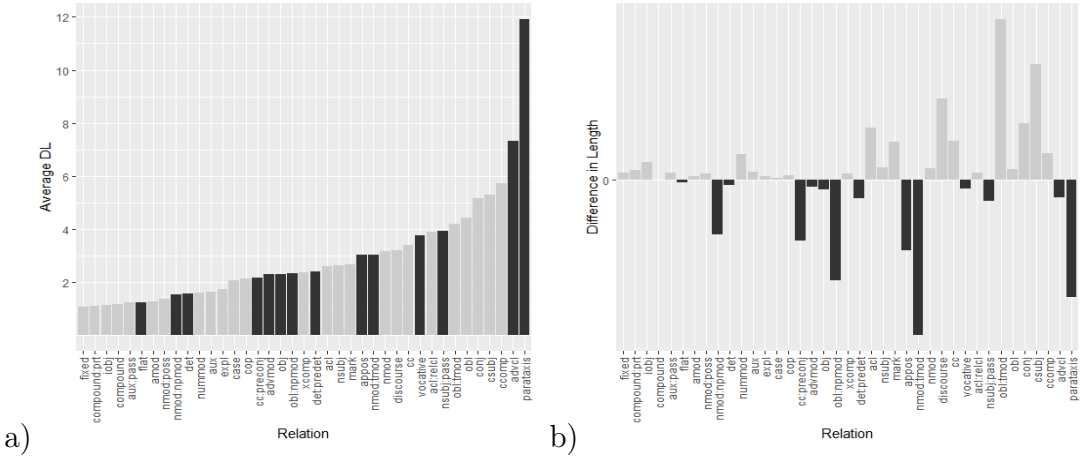
Figures 6.28, 6.29 and 6.40 show the results for this comparison[12]. As above, the charts on the left (Figures a) visualise average dependency length of each relation, sorted according to increasing average dependency length. In contrast to the visualisations above, the colours depict differences in average relation length: grey bars depict longer average relation length in originals, black refers to longer average relations length in interpreting. Figures b) use the same order of relations, but visualise length differences for each relation in originals vs. interpreting. As in the previous visualisations, the length of the bars - regardless of depicted in the negative or

[12]The extracted average DL and differences in DL used as basis for these visualisations are given in the Appendix (Tables A.16 - A.18).



a) b)

Figure 6.28: Dependency relations and their average dependency length (a) and their differences in length (b) for ORG EN vs. SI DE EN. Grey denotes longer relation length in original, black marks longer relation length in simultaneous interpreting. Normalised to the number of relations.



a) b)

Figure 6.29: Dependency relations and their average dependency length (a) and their differences in length (b) for ORG EN vs. SI ES EN. Grey denotes longer relation length in original, black marks longer relation length in simultaneous interpreting. Normalised to the number of relations.

positive range - refer to differences in average length in each subset. The basis of normalisation used for this comparison is the number of relations.

For the comparable English dataset, the following observations can be made. *parataxis* as the longest relation on average is notably longer in interpreting from both source languages (German and Spanish) than in comparable English originals. As stated in the previous section, this relation is also used more frequently in interpreting. However, it mainly occurs in the context of incorrect parsing output due to filled pauses. Therefore, the results for *parataxis* are unreliable and are not considered

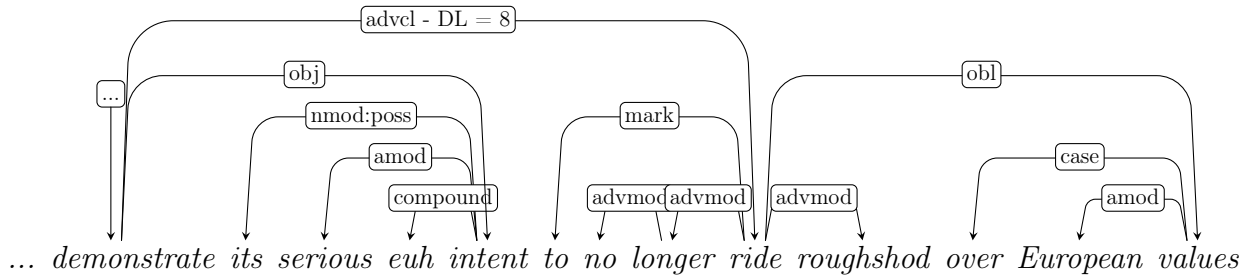


Figure 6.30: Example for relation *advcl*.

further.

advcl (adverbial clause modifier) is the second-longest relation on average for both English comparable datasets. Although *advcl* occurs more frequently in English originals overall (cf. previous section), when this relation is used in interpreting, the distance to its head tends to be longer than in comparable originals.

ORG DE	SI DE EN
aber wir sind der Meinung dass hier viel getan werden muss dass die türkische Regierung ihren ernsthaften Willen demonstrieren muss europäische Grundrechte nicht mehr zu missachten (*xcomp*) wie sie das im Moment tut ↲	but we think a great deal needs to be done if euh Turkey is to really demonstrate its serious euh intent to no longer ride (*advcl*) roughshod over European values ↲

Table 6.30: Example sentence for *advcl* (cf. Figure 6.30).

The example in Table 6.30 shows a long and complex sentence in the German source, which is not split, but rather its hypotactic structure is maintained in the interpreter's output. The clause of the source, *wie sie das im Moment tut*, is omitted in the target, which renders the target less complex. Furthermore, the clausal relation of the subordination differs from source (consecutive clause) to target (conditional clause), but in both cases the clausal relations (*xcomp* and *advcl)* refer to an infinitive used in the subordinate clause with a large distance from its head (source: distance to demonstrieren = 7 tokens; target: distance to demonstrate = 8 tokens). This long dependency distance of *advcl* may therefore be the result of maintaining the structure if the source language (shining-through).

Furthermore, two types of *nominal modifiers* tend to have a longer distance to their heads in interpreting than in originals. This is again true for both comparable English datasets. The difference is particularly large for *nmod:tmod* (temporal modifier) with an average dependency length of 2.74 in ORG EN, 4.33 in SI DE EN and 3.86 in SI ES EN, respectively. The differences are not as large but still noticeable for *nmod:npmod*

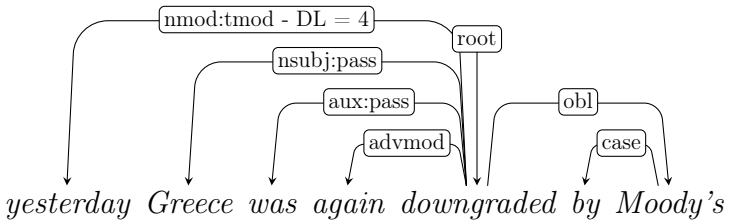


Figure 6.31: Example for relation *nmod:tmod*.

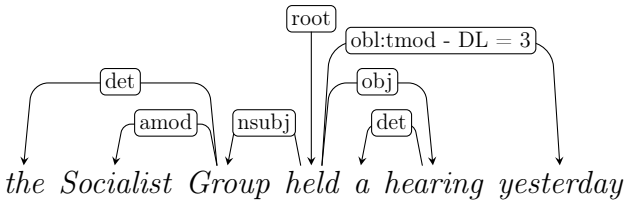


Figure 6.32: Example for relation *obl:tmod*.

(noun phrase as adverbial modifier) with av DL = 1.42 in ORG EN, av DL = 1.81 in SI DE EN and av DL = 1.82 in SI ES EN (cf. Figures 6.28 and 6.29).

When analysing the examples for *npmod:tmod*, it has to be considered that they occur very infrequently overall (22 hits in the ORG EN/SI DE EN subset and 26 hits in ORG EN/SI ES EN). This relation is used for nominal temporal modifiers that are placed at the beginning of the sentence (cf. example in Figure 6.31). If the same nominal modifier is used at the end of a sentence, it is grouped as an oblique temporal modifier (cf. the example in Figure 6.32). Therefore, the two relations should be looked at together.

As stated above, *nmod:tmod* shows longer distances to its head (in this case the *root* of the sentences) in English interpreting from German as well as from Spanish compared to English originals. In the analysis of relations' frequency (previous section), no clear trend was observed. *nmod:tmod* was found to be more frequent in English originals when compared to interpreting from German, but less frequent in English originals when compared to interpreting from Spanish. These frequency differences can be explained again by the influence of the source speech. Looking at examples for this relation in interpreting and the source sentences on which they are based, it can be stated that the position of temporal modification in the source is also mostly at the beginning of the sentence (cf. Figure 6.33). Interpreters maintain this position in their target output. Shifting the source input (*gestern Abend*) to the end of the sentence, which would be a possible alternative position in English, would shorten the distance from DL = 4 as in the example to DL = 3 (cf. Figure 6.33 vs. Figure 6.34). However, this alternative only presents an optimised position for the temporal reference from a dependency length perspective in the target and does not consider the translation process where a source element that was uttered by the source speaker needs to be maintained in memory until it is produced by the interpreter and

the dependency relation in the target is resolved. Restructuring the source input and producing a shorter dependency distance in the target would potentially result in higher memory costs for storing and restructuring the source input. Figure 6.33 shows that by maintaining the initial sentences position for the temporal reference results in the shortest possible storage costs for *gestern - yesterday*[13]. Furthermore, it can be noted that overall long temporal modification at the end of a sentence (resulting in the relation *obl:tmod*) tends to be avoided in interpreting since this relation is longer in English originals overall (cf. Figures 6.28 and 6.29).

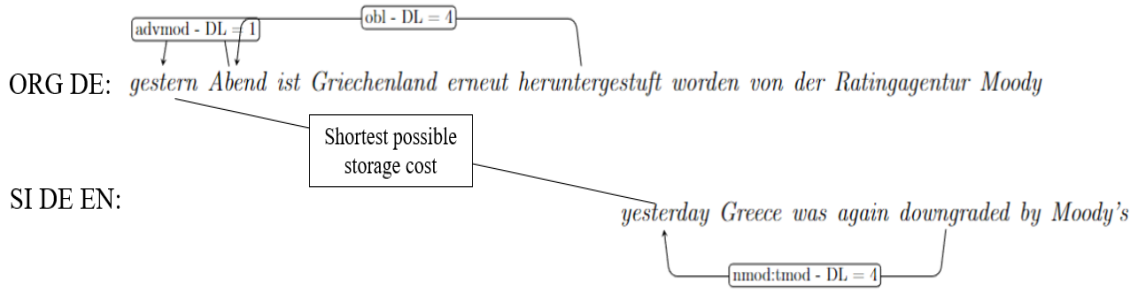


Figure 6.33: Example sentence for temporal relation in source ORG DE and target SI DE EN - maintaining position of source with longer dependency relation in target.

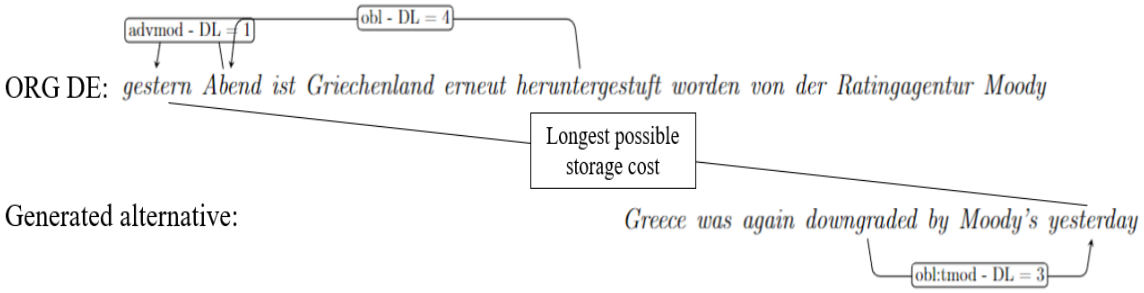


Figure 6.34: Example sentence for temporal relation in source ORG DE and generated alternative in target - shorter dependency relation in target leading to higher storage costs for source input.

Temporal modification, such as the sentence initial modification *nmod:tmod*, or *obl:tmod* in sentence final position in the English target, as well as the *obl* temporal modification in the German source, are elements that are directly linked to the sentence root or the head of the clause. However, as they do not require modification, such as inflection or reordering of sentence elements in English, longer distances to their head may not require as much memory effort in production as for sentence

[13]Additionally, the specific time reference *Abend* as well as nominal modification *Ratingagentur* are omitted in the target, which leads to shorter the dependency relations in the interpreter's output compared to an output including these elements.

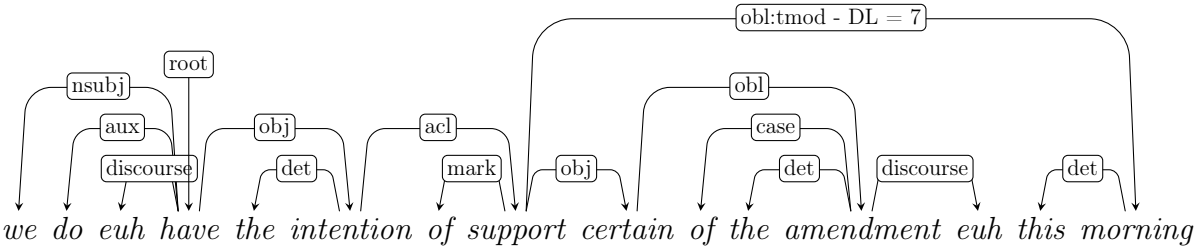


Figure 6.35: Example for relation *obl:tmod* in ORG EN.

elements that require such modification effort. Therefore, minimising dependency length may not be as crucial for such temporal modifiers and higher benefit in memory costs for the interpreter may be gained by reproducing the source input as soon as syntactically possible in the target (in this case in sentence initial position) instead of storing it until they are produced in an alternative position (such as in *Greece was again downgraded yesterday by Moody's* or as in Figure 6.34).

The alternative position for such temporal modification placed within or at the end of a clause is labelled as *obl:tmod*. These temporal modifiers and other non-core oblique arguments or adjuncts *obl* are found to be longer in English originals than interpreting from both different source languages (cf. Figures 6.28 and 6.29). They were also found to be more frequent in English originals than interpreting (cf. Figures 6.7 and 6.8).

Looking at examples in the English originals subset shows that English original speakers tend to produce the temporal modification at the end of a clause or sentence. This can lead to fairly long dependency relations, as in the example in Figure 6.35. Interpreters, on the other hand, are influenced by the order in which the source input is structured. German original speeches are fairly flexible in the position of temporal modification. Figure 6.36 shows that the interpreter produces the source input in a very linear way in the target (*nachdem - after, Ministerpräsident Topolánek - Prime Minister Topolánek, heute - this morning* etc.). The alternative position for the temporal reference at the end of the clause (cf. the alternative generated in Figure 6.37) is avoided by the interpreters. It would lead to longer dependency relations and require additional effort to restructure and intermediately store the elements. Maintaining the source input sequence can be seen as the most efficient production mode for the interpreter, which also leads to the most efficient dependency distance for *oblique temporal modifiers*.

The relation *clausal subject* as clausal syntactic subject of a clause (*csubj*) is a further dependency relation that is clearly longer in English originals compared to English interpreting from both source languages. This relation is generally linked to the root of the sentence, mostly the main lexical verb or other predicate of the subject clause. Furthermore, the relation is also used for clausal subjects in cleft sentences

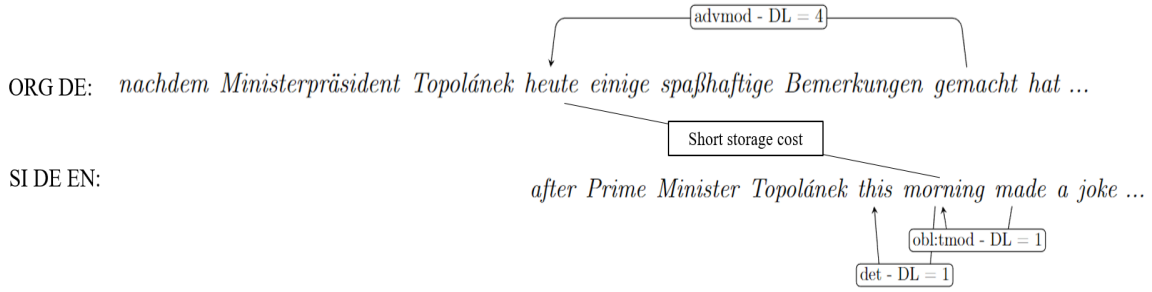


Figure 6.36: Example sentence for temporal relation in source ORG DE and target SI DE EN - maintaining position of source with shorter dependency relation in target.

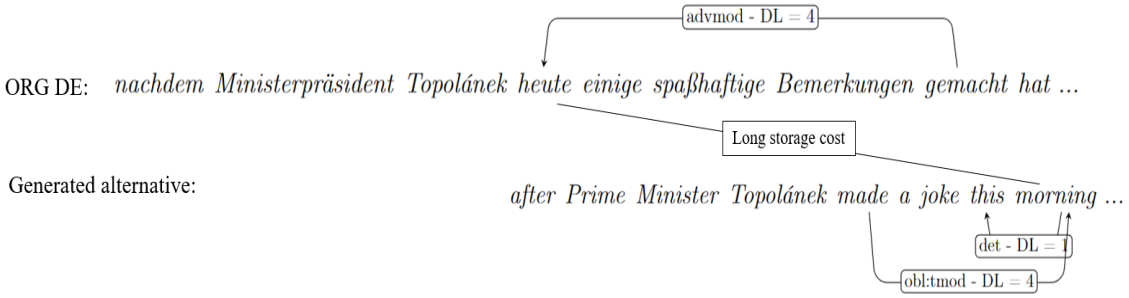


Figure 6.37: Example sentence for temporal relation in source ORG DE and generated alternative in target - longer dependency relation in target and higher storage costs for source input.

(de Marneffe et al., 2014).

Although no clear frequency differences were found in the use of this relation between the different subsets in the analysis in the previous section, comparing the dependency length differences shows a longer use of *csubj* in original than in interpreting. An explanation can be found in multiple examples. On the one hand, long clausal subjects tend to be avoided altogether in interpreting. The example in Figure 6.38 and Table 6.31 shows a long *csubj* with DL = 8 in the English original speech that is replaced with a pronominal subject and modal verb in interpreting. Although this is clearly a shift in the pragmatic function, the interpreted version is shorter and less complex. On the other hand, interpreters tend to split long and complex source sentences (cf. Section 6.2.2), where long *csubj* are likely to occur. This is shown in the example in Figure 6.39, where the long and complex source input leads to a long distance of *csubj* to the head (*clear* - DL = 15). This is split into several independent clauses in interpreting, preventing such long relations altogether (cf. Table 6.32).

Finally, clear source language effects shining-through can also be observed when comparing average relation length. The relation *conj* tends to be longer in interpreting from German when compared to English originals, but shorter in interpreting from

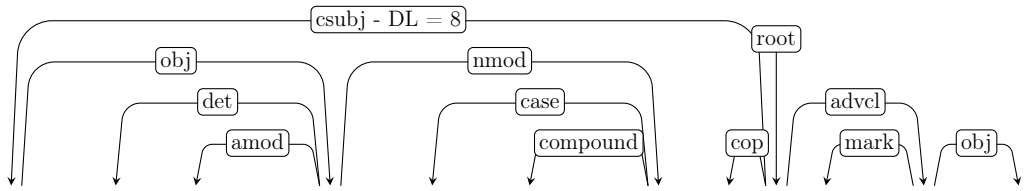


Figure 6.38: Source speech for the example for relation *csubj*.

ORG EN	SI EN DE
admitting (*csubj*) the central importance of population growth is key to addressing it ↲	wir müssen die Bedeutung des Bevölkerungswachstums anerkennen ↲

Table 6.31: Example sentence for *csubj* and interpreting (cf. Figure 6.38).

Spanish when compared to English originals.

For the English dataset, it can therefore be summarised that generally longer dependency relations, except for *advcl*, tend to be longer in originals. Especially *clausal subjects* tend to split and resolved in other ways in interpreting. For other relations, especially *temporal modifications* realised as *nmod:tmod* or *obl:tmod*, the source language shinning through can be observed.

For the German comparable dataset, it can be observed that almost all dependency relations tend to be longer in originals than interpreting from English as source (Figure 6.40). Especially, the longest relations are notably longer in originals than in interpreting. German's fairly flexible word order in combination with interpreters' tendency to produce shorter sentences seems to allow for source language input from English to be structured in a way that optimises dependency length of individual dependency relations in German altogether.

In particular, long clausal relations (*ccomp, xcomp, csubj, csubj:pass* and *advcl*) tend to be longer in German originals than in German interpreting. This observation

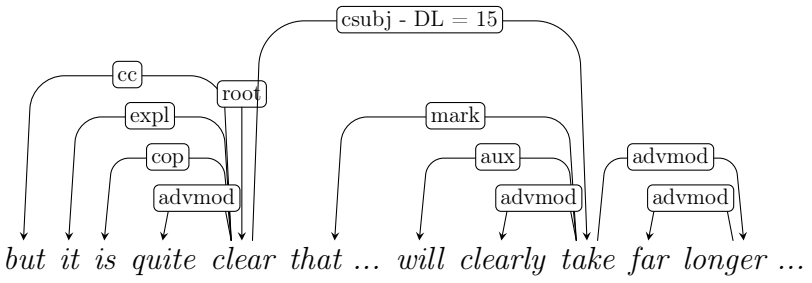
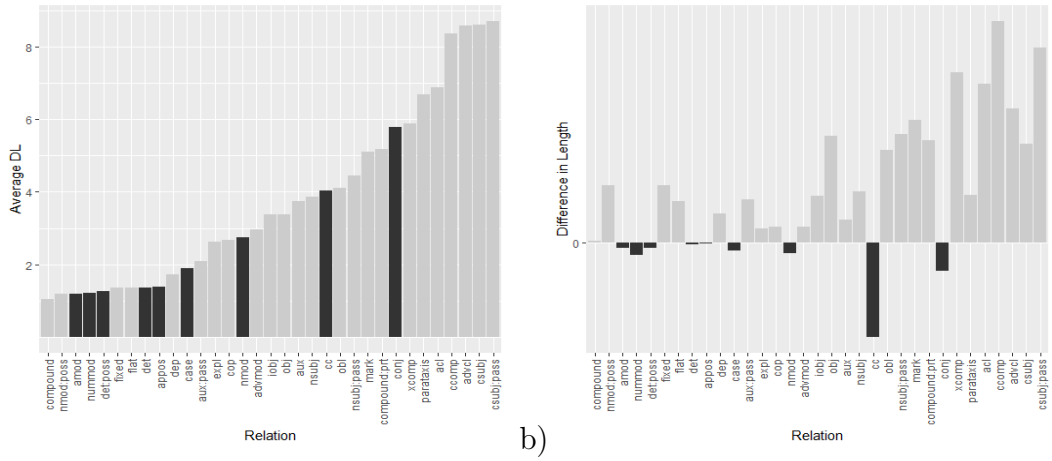


Figure 6.39: Source speech for the example for relation *csubj*, split in interpreting.

ORG EN	SI EN DE
but it's quite clear that euh solutions that may involve euh some new European-level judicial mechanisms will clearly take (*csubj*) far longer and and potentially be more controversial than picking up some of the alternative dispute resolution measures or also euh using the existing consumer cooperation measures that have been put in place ↲	also es ist angemessen dass man nach Lösungen sucht die euh zum Teil auf europäischer Ebene neue Rechtstraditionen bedeuten werden ↲
	und das wird natürlich länger dauern und wird umstritten sein ↲
	wenn man da jetzt alternative Rechtsschlichtungsverfahren miteinbezieht euh euh dann kann man auch sehen dass man Rechtsdurchsetzungsverfahren berücksichtigen muss die schon zum Teil gelten ↲

Table 6.32: Example sentence for *csubj*, split in interpreting.



a) b)

Figure 6.40: Dependency relations and their average dependency length (a) and their differences in length (b) for ORG DE vs. SI EN DE. Grey denotes longer relation length in original, black marks longer relation length in simultaneous interpreting. Normalised to the number of relations.

is in line with the results from the English comparable datasets, where *csubj* was also found to be longer in English originals than in interpreting into English from German or Spanish.

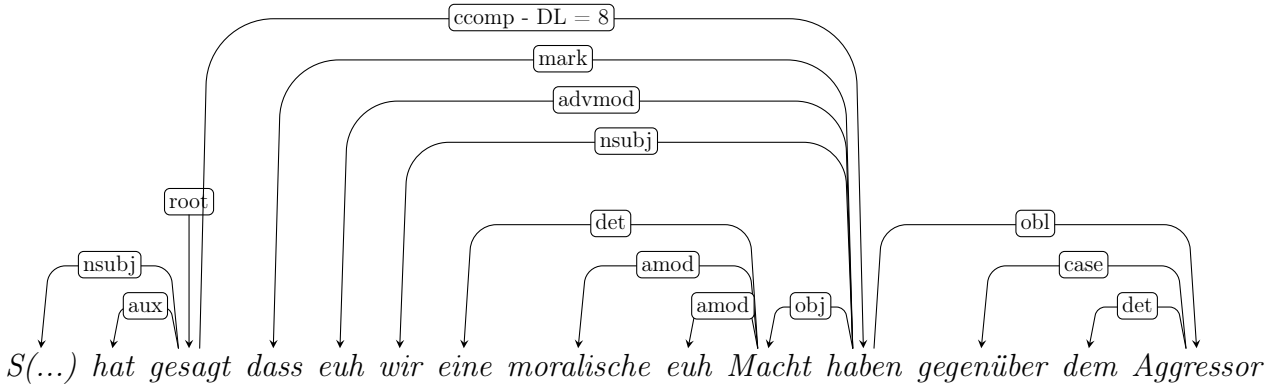


Figure 6.41: Example for relation *ccomp* and extraposition in interpreting.

Figure 6.41 shows an example for *ccomp* from German interpreting. *Clausal complements* are used for clausal dependents of verbs that function like a complement of that verb, typically in German subordinates. *ccomp* is also used for reported speech (de Marneffe et al., 2014). German subordinate clauses typically use an SOV (Subject-Object-Verb) order, with the finite verb produced in final position. Although the standard word order in German subordinates is SOV (Gerritsen and Stein, 1992; Eisenberg, 1994), German allows extraposing syntactic elements after the verb group. Collard et al. (2019) show that German and Dutch interpreters tend to produce verbs in subordinate clauses earlier and extrapose content in the afterfield significantly more often than original speakers of the same language. Therefore, interpreters tend to produce subordinates as in Figure 6.41, in this case with a DL = 8, while original speakers tend to produce the verb group in the final position as in the generated alternative for the same sentence in Figure 6.42 with a longer dependency distance for *ccomp* of DL = 11. This may have to do with interpreters' tendency to reproduce source input in a more linear fashion, with source content that was uttered last in the source also being produced last in interpreting, but also with optimised management of cognitive resources and avoiding long open dependency relations as they would have to be maintained in memory for a longer period of time. However, it also has to be noted that the sentence length was found to be significantly shorter in interpreting than in original speech (cf. Section 6.2.1). This can be seen as another explanation for long dependency relations occurring more frequently in originals.

The only clear exception to the trend of dependency relations being longer on average in German originals than in German interpreting is linked to coordination. *cc (coordination conjunction)* and *conj (conjunct)* are longer in interpreting than originals (cf. Figure 6.40). *cc* is the relation between a conjunct and a preceding coordinating conjunction, and *conj* refers to the relation between two elements connected by a coordinating conjunction (de Marneffe et al., 2014).

Coordinating and subordinating conjunctions *aber, dann, denn, da* were found to be distinctive for German interpreting when compared to German originals (cf.

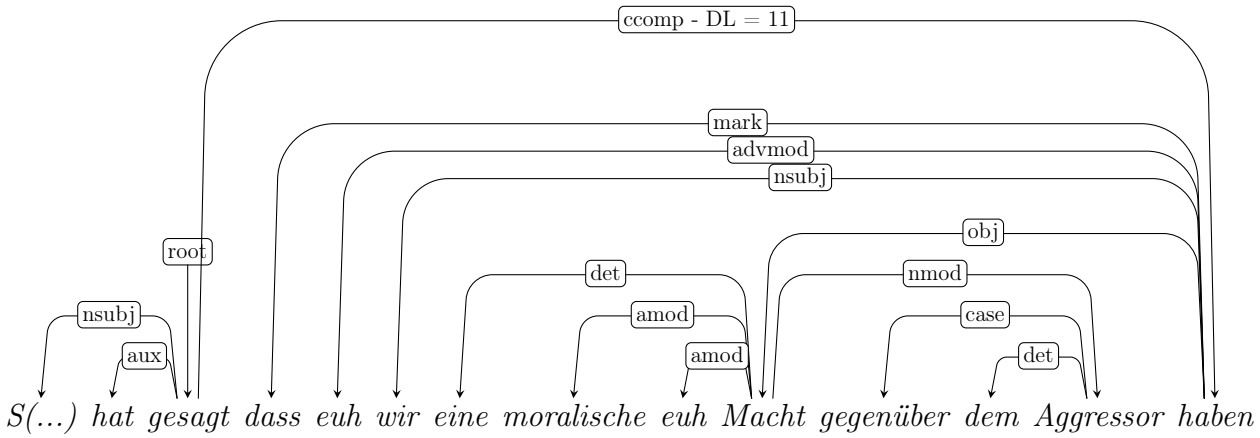


Figure 6.42: Generated alternative for relation *ccomp* and without extraposition.

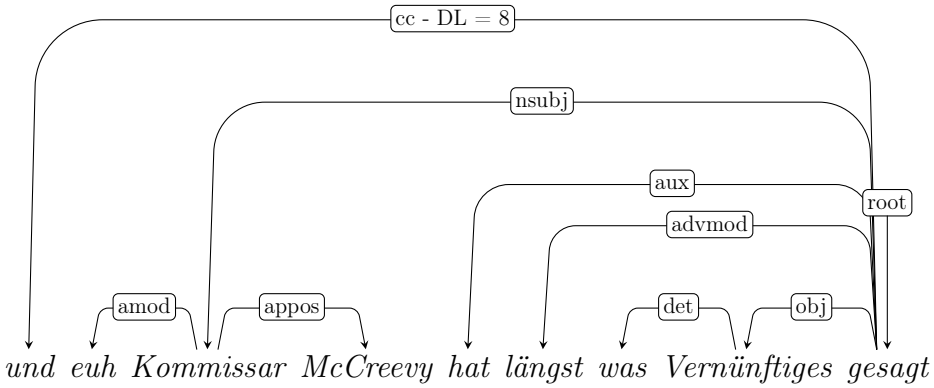


Figure 6.43: Example for relation *cc* in interpreting.

Section 4.2). Furthermore, as seen in Section 6.2.2, interpreters tend to split the source segments into several independent target segments. However, these split sentences tend to be linked with a coordinating conjunction: 29.2% of all sentences in German interpreting are initiated with a coordinating conjunction (1189 out of 4076) vs. 21.5% of all sentences in German originals (733 out of 3409). In German, these coordinating conjunctions at the beginning of the sentence tend to have fairly long distances from their heads (generally the root of the sentence) as in the example in Figure 6.43. The higher frequency of coordinating conjunctions at the beginning of the sentence therefore explains the relation *coordinating conjunction* being longer on average in German interpreting than in originals. These longer dependency distances for the relation *cc* can be interpreted as means of cognitive management by splitting source sentences rather than adding cognitive strain by creating longer dependency distances.

The relation *conj (conjunct)* is the relation between two elements connected by a coordinating conjunction. This can refer to coordination below the clause level (*apples and pears (conj)*) or, as is the case for longer *conj* relations, to coordination on clause level. As the segmentation guidelines for EPIC UdS V3 state that independent

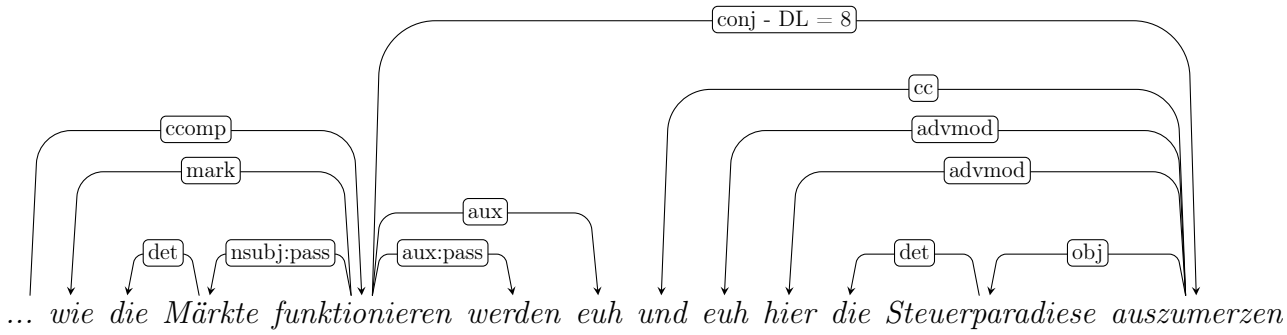


Figure 6.44: Example for relation *conj* in interpreting.

coordinated clauses are supposed to be segmented separately, *conj* is supposed to relate to coordination of dependent clauses as in Figure 6.44.

As shown in Section 4.2 and by Gast and Borges (2023), interpreting tends to use a more verbal style rather than nominal expressions. These clausal relations may result in longer dependency distances compared to a nominal expression for the same input. This trend may be reinforced by the source language influence with German interpreting based on English sources, as English sources tend to be more verbal than German in general. However, longer clausal relations in German interpreting are only observed for the relation *conj*. An inspection of the *conj* corpus examples reveals some inconsistencies in the segmentation of independent clauses. Segmentation is based on manual annotation which can include some human error. More importantly, corpus analysis shows that this relation was frequently not tagged correctly, e.g. it was used to tag a relative clause, but as the comma before the relative pronoun was not transcribed, the whole clause was not recognised correctly. Therefore, the overall result of *conj* being longer on average in German interpreting than German originals cannot be considered reliable.

For the German subsets overall, it can be concluded that German's free word order as well as interpreters tendency to split source sentences and produce shorter sentences altogether leads to shorter, more optimised dependency relations in interpreting. However, the source language input structure is sometimes reflected in the target output, which may lead to longer dependency relations in some cases (temporal modifiers) and reduce dependency length in others (clausal complements). Therefore, a shorter average length of individual dependency relations in German interpreting overall makes the interpreting product simpler in general. Interpreters need to maintain fewer items in memory to close a relation than comparable German original speakers.

To summarise the analysis of average length of dependency relations, it can be concluded that the tendency to minimise dependency length as an indicator of cognitive resources management is greater in German than in English. This is in line

with findings of Sections 6.2.1 and 6.2.2, which show a greater tendency to split the source sentence into several independent target sentences in German than English (20.75% of English source sentences split into multiple independent target sentences in German interpreting vs. 16.45% of German source sentences split into multiple target sentences in English interpreting). The shorter sentence length in interpreting as a result of splitting (and sometimes omitting source content) can partly explain these shorter dependency relations in interpreting. However, for some individual relations interesting trends can be observed such as *clausal subject*, which tends to be avoided or shorter in English interpreting. Instead *adverbial clause modifiers* are used and tend to be longer in English interpreting. The use of clausal modification as subordinate clauses may also be seen as segmenting the source input into information chunks, however, with dependency relations that still have to be maintained in memory until the relation is closed. Some of these effects may result from source language shining-through, as observed for *temporal modifications*. This may lead to longer dependency relations, however, this is counterweighted by interpreters only having to store source input for the least amount of time in memory and no effort for restructuring the source input has to be made. The effect can especially be seen for *clausal complements* in German, where interpreters in German produce the verb group in the subordinate first before extra-posing further syntactic elements (Collard et al., 2019).

Therefore, it can be concluded that interpreters simplify by generally producing shorter dependency relations, especially for relations with high average dependency length, but also manage cognitive resources by maintaining source input in a fairly linear way and by producing shorter sentences overall.

6.3 Summary of results on syntactic simplification

In a first general syntactic analysis, sentence length was investigated on token basis (Section 6.2.1). The analysis revealed that interpreters produce shorter sentences than comparable original speakers, as well as shorter sentences than used in the source speeches on which interpreting was based. This is in line with trends previously observed in the literature for other languages (Bernardini et al., 2016; Liu et al., 2023).

The main findings of the alignment analysis (Section 6.2.2) can be summarised as follows. Interpreters generally produce one target sentence for one source segment and maintain the overall macrostructure of the source speech. This is in line with Liontou (2011), who finds that interpreters mostly do not rearrange source input and goes against Goldman-Eisler's (1971) observation of interpreters imposing their own segmentation, although the current study does not go below the sentence level and cannot make claims about rearrangement at chunk level.

Interestingly, the second most frequent category in the alignment analysis is inter-

preters splitting source sentences into several target sentences, which was defined by Baker (1996) as a means of simplification. This trend was also reported in the literature as a source text conditioned interpreting strategy (Kalina, 1998), used at the micro level to reduce cognitive load. By producing several shorter complete sentences, source speech elements do not need to be stored in memory as long as compared to reproducing a long and complex source speech sentence. Sentence splitting in interpreting was also previously observed in the literature, indicated by general means such as more connectives, more demonstratives, and more sentences in a text chunk (He et al., 2016) or as qualitative result of a word class analysis (Gast and Borges, 2023). This study is the first to our knowledge to show such a result in a manual alignment analysis of the entire German to English and English to German interpreting corpus. With regard to simplification, it can be stated that by breaking up longer sentences into several shorter ones, syntactic complexity is reduced, and a lower level of embeddedness is generally defined as more simplified (cf. Section 2.5.2). Furthermore, it was observed that these multiple target sentences sometimes use the same or similar sentence structure in subsequent sentences. This syntactic repetition could also be interpreted as more simplified. When splitting longer sentences, it was also observed that interpreters occasionally left out repetitions or modifiers. Such omissions can be strategic *selection* and *deletion* where redundant or less important components of the source speech are not rendered in the target speech (Kalina, 1998) and are sometimes also referred to in the context of cognitively challenging conditions (Gile, 1995, 1999; Setton, 1999). Although this can be rather linked to the lexical aspects discussed in Chapter 5, it should be mentioned here that, in addition to making the target output shorter on the syntactic level, such omissions of lexical words would also make the target output less dense and contribute to a more simplified output.

The third most frequent alignment category are none-alignments. This mostly includes omitted target sentences of source speech sentences at the end of a speech (such as "Thank you"). These sentences are generally very short and are expected at the end of a speech. Leaving them out in the target speech can be seen as rendering the target speech less simplified. Such additions of "Thank you" or "this is what we need to do" by interpreters at the end of their speech are rare but do occur occasionally. Such additions of expected items would then be seen as simplification of the target. The effect of other sentence omissions where larger content is left out in interpreting depends on whether the omitted content is required to establish cohesion and cannot be assessed in the context of this study.

The least frequent category in the alignment analysis is the category n:1 where interpreters merge several source sentences into one target sentence. The literature refers to this where the length of an utterance is compressed while preserving the semantic content of a message as *filtering* (Al-Khanji et al., 2000) or *summarising* (Li, 2015). It is the least frequent alignment type in our data, while other studies find such summarising to occur more frequently when the focus of investigation is particularly on strategies employed by interpreters (Al-Khanji et al., 2000). At word level, such

merged sentences can be longer and more complex than individual source sentences on which they are based. However, it is the least frequent alignment category in our dataset, and these few cases of longer target sentences do not counterweight the general observed trend of shorter sentences in interpreting, observed in Section 6.2.1.

These observed trends can generally be related to syntactic simplification in interpreting and confirm **RQ 3a**. Syntactic simplification in interpreting can be observed via a sentence length analysis (comparable and parallel approach) as well as the alignment analysis (parallel approach).

The main part of this chapter (Section 6.2.3) focused on a dependency length analysis and the hypothesis that interpreters simplify their output syntactically by minimising distances of dependent relations. However, **RQ 3b**, asking whether syntactic simplification measured via dependency length minimisation can be observed in the interpreting product, can generally not be confirmed. In a comparable approach, comparing originals and interpreting in the same language, no general simplification effect via shorter dependency length, but rather a source speech influence was observed. Although interpreting using shorter sentences overall (Section 6.2.1), a clear trend of reduced syntactic complexity in interpreting was not found. While dependency length minimisation in interpreting, investigated via average dependency length for all dependency relations, was not confirmed, only a tendency of reduced syntactic complexity, measured via average maximum dependency length was observed. However, this may be explained by the shorter sentence length in interpreting overall. The percentage of very short dependency relations (DL =1) in the different subsets also did not show signs of simplification, and an analysis of short vs long sentences only showed a slight trend of long sentences in German interpreting to use shorter dependency relations than in German originals.

With no general dependency length minimisation effect found in interpreting and shining-through of source language structure observed via a comparable approach, **RQ 3c** on the triggers of a potential minimisation effect can be answered as follows. Despite the general rejection of DLM in interpreting, the qualitative analysis of individual dependency relations, looking at their frequency and average length in the individual subsets, showed interesting traces of the differences observed in the different subsets. This includes some dependency length minimisation trends. For example, the analysis of the average length of dependency relations showed that interpreters generally produce shorter dependency relations than original speakers, especially for relations with high average dependency length. Especially long clausal dependency relations seem to be avoided by the interpreters. Furthermore, long dependency relations were linked to originals, with long relations occurring more frequently in English originals, as well as long relations shown to be longer in German originals. However, such length minimisation trends for individual dependency relations quantitatively do not result in DLM overall.

The present analysis rather showed that interpreters tend to produce source input in a linear way and as early as (syntactically) possible in the target, which sometimes may lead to longer dependency relations in interpreting (e.g. *temporal modifiers* in English or *clausal complements* in German). The observed pattern of interpreters splitting longer source sentences into several coordinated sentences, resulting in some longer or more frequency dependency relations (e.g. coordinating conjunctions or clausal relations), also contribute to the fact that no general DLM was found in interpreting. Generally, shining-through of source language structure, which was also observed via a comparable approach, was also observed via a parallel analysis. The source language effect combined with other interpreting process related effects counterbalance a possible dependency relation minimisation effect in interpreting (**RQ 3c**).

In addition to answering the research questions laid out to investigate in this syntactic investigation, the analysis of dependency relation frequency also confirmed the results obtained in the exploitative investigations (Section 4.2) with interpreting using more characteristics of spoken language such as intensifiers, discourse markers, filled pauses, and auxiliaries. These were possibly linked to omissions and additions in interpreting. Furthermore, the analysis of individual relations confirmed the observations from the alignment analysis (Section 6.2.2) as a result of splitting sentences. Additionally, traces of the interpreting strategies summarising and omissions, as discussed in the background Section 2.2.2, were also observed.

The study also revealed some clear language differences. It can be particularly noted that the average dependency length for almost all individual dependency relations in German interpreting was lower than for German originals. This is in line with a stronger tendency of splitting English source input in German interpreting compared to the reverse interpreting direction (German originals interpreted into English). Another possible explanation for these differences is also the freer word order in German allowing for more flexibility in the sentence structure and, therefore, more flexibility for minimisation of dependency distances of individual relations. However, in the literature, the possibility of DLM was found to a higher degree in English than German (Gildea and Temperley, 2010; Futrell et al., 2015; Dyer, 2023).

On a more general note, the study results revealed some limitations of the approach used. It was shown that using DLM as approximation of production costs in interpreting only reflects part of the interpreter's memory load. Source input has to be maintained in the interpreter's memory until it is produced in the target and the dependency relation is resolved in the target. Using dependency length measures generated independently for source and target speech does not fully capture syntactic processing in interpreting. To cover the entire interpreting process, a bilingual syntactic processing model would be required, which considers memory load for production while at the same time allowing processing of syntactic elements of the source

speech to be considered. While more advanced models reflect memory use in sentence processing, e.g. combining the information-theoretic notion of surprisal with the cost of storing some amount of information about past context (memory-surprisal tradeoff (Hahn et al., 2021)). These models only reflect monolingual processes and do not consider particular characteristics of the interpreting process with bilingual processing, including source text processing in one language and target text production in a second language, and particularly time components such as EVS and the implications for comprehension, production, and intermediately storing of source elements that need to be produced in the target speech.

In addition to a more sophisticated model, the corpus data used in such a model would also require further annotation. The EPIC-UdS corpus annotation includes only sentence alignment of source and target. More fine grained annotation of source and target word time tags as well as source and target alignment below the sentence level would be required. This study showed qualitative evidence of interpreters maintaining fairly linear order of source input in the target. In order to fully investigate this as potential means for cognitive relief, analysing aligned content words of source and target would shed more light on this phenomenon.

To state a further limitation of the approach used in this study, it should be mentioned that the corpus version and individual annotation decisions may also influence dependency length results. The decision was taken to remain as true to the original spoken product as possible for the corpus version EPIC UdS V3. While some manual corrections of some POS before parsing were carried out, filled pauses and other typical spoken features are still included in this corpus version. The initial assessment of parsing accuracy for this corpus version was very promising (cf. Section 3.1.4). However, the qualitative analysis of individual dependency relations in Section 6.2.3 showed that some dependency relations are more affected by these spoken elements and missing mid-sentence punctuation than others (cf. results for *parataxis* in the English dataset for parsing errors due to false starts and *conj* in the German dataset for errors due to missing mid-sentence punctuation).

From a perception perspective, the assumption that linguistic items have to be maintained in memory until the dependency relation is closed does not fully apply to filled pauses. Including them as intervening words in the corpus distorts the counts of dependency distances, especially as they occur more frequently in interpreting than originals. Furthermore, by adding mid-sentence punctuation, syntactic parsing accuracy could be improved further, which is available for later EPIC-UdS corpus variants. This is particularly relevant for German.

Despite these minor shortcomings, the study showed clear signs of syntactic simplification. However, this was achieved mainly by interpreters producing shorter sentences. Syntactic simplification via dependency length minimisation was observed only for individual relations but cannot be seen as the main strategy for interpreters to optimise their cognitive resources. Producing shorter sentences by splitting the

source input and maintaining source input in a fairly linear way was found to be a more prominent pattern.

Part IV

Summary and conclusion

Chapter 7

Main findings and conclusion

7.1 Combined results and discussion

This thesis aimed to give a more detailed understanding of the specific properties of interpreted language, also referred to as *interpretese*, and investigated the idea that, due to the particular constraints of the interpreting process and resulting implications for predictability and memory optimisation, interpretese would be characterised by simplification, observable via lexical and syntactic choices made by the interpreters. The main findings can be summarised as follows. Interpretese was found to be primarily characterised by features of spoken language such as filled pauses, intensifiers, deictics, discourse makers, and verbal instead of nominal use. These findings are the main results of the exploratory analysis in this thesis (Chapter 4), where interpretese was investigated via relative entropy (Kullback-Leibler divergence, KLD). The findings are in line with the literature that states that interpreted language is particularly characterised by spoken language features (Shlesinger and Ordan, 2012; Karakanta et al., 2021; Gast and Borges, 2023). These results were also confirmed by the investigations of the qualitative analyses in this thesis, investigating the lexical and syntactic properties of interpreting via surprisal and Dependency Locality Theory (DLT) (Chapters 5 and 6). Furthermore, analysing different delivery modes in the lexical studies also confirmed a trend of interpreting being particularly spoken.

The main focus of this thesis was to investigate simplification as one potential characteristic of interpretese, which is also one of the known properties of written translations. Hypothesising that the interpreting process would force interpreters to opt for lexical and syntactic choices that are beneficial for processing and memory, which are traceable as the result of simplification in the interpreting product, it was investigated whether interpreters use more expected words in context, i.e. words with lower surprisal (Chapter 5) and opt for short dependency distances (Chapter 6). The lexical investigations were complemented by measures of simplification that are tra-

ditionally used in corpus-based studies, such as lexical density, lexical sophistication, and lexical variation. The studies on simplification used a comparable and parallel study design, comparing interpreted language to original spoken language in the same language in the former and source speeches to target speeches in the latter approach.

The results for simplification can be summarised as follows. Using a comparable approach, simplification was observed at the lexical and syntactic level. At the lexical level, simplification was observed using traditional measures (Section 5.1). Lexical simplification was evident through lower lexical density overall for interpreting compared to original speeches in the same language, as well as for only impromptu delivery and for speeches delivered at the same delivery rate. Simplification was also found through lower lexical sophistication and lower lexical variation in interpreting, as well as through shorter encodings at the word and noun phrase level. Approximating simplification using information theory (Section 5.2) revealed that simplification was generally observed through the use of content words with lower surprisal in interpreting. This was the case when investigating all tokens, verbal POS (except in impromptu delivery mode in German), and nouns in English (not in German). Furthermore, low surprisal verbs were generally more frequent in interpreting than in originals in both languages, as well as low surprisal nouns were found more frequent in English interpreting. However, low surprisal nouns were more frequent in German originals than in German interpreting. High surprisal verbs and nouns, on the other hand, were used more frequently in originals than in interpreting in both languages. The results show that opting for more standardised and expected translation solutions is a mechanism employed by interpreters when dealing with lower surprisal and expected input. For unexpected (nominal) input, no frequently used translation equivalent is available for the interpreter and interpreters may be influenced by the source language input.

Using traditional measures for the lexical investigations, the observed simplification trends were more pronounced for German interpreting compared to German originals using traditional measures, but the same trend was also clearly visible for the English comparable dataset. When using surprisal to approximate simplification, the observed trend was stronger in English and rather for verbal POS, which are more characteristic of interpreting. A potential explanation for this difference was found in the way surprisal was modelled in the corpus, linked to small data size and influenced by German's richer morphology and compounding. Furthermore, the comparative studies showed an influence on the simplification patterns of delivery mode and preparedness of the original speeches. Generally, these results on lexical simplification show that the interpreting product is simpler than comparable original speeches. This was shown for both languages investigated, for German as well as for interpreted English from two source languages.

The measure of surprisal reflects, at least to some extent, lexical variation in context. In that sense, lexical variation as studied with traditional measures and surprisal of lexical words as expectedness in context are linked. Therefore, the results for the

surprisal studies complement the results of lower lexical variation in interpreting as found in the studies using traditional measures in this thesis, and are in line with the results described in the literature (Russo et al., 2012; Kajzer-Wietrzny and Ivaska, 2020; Xu and Li, 2022).

Using a comparable approach at the syntactic level gives slightly different results. In general, sentences in interpreting are shorter than in comparable originals (Section 6.2.1), which is consistent with trends previously stated in the literature for other languages (Bernardini et al., 2016; Liu et al., 2023) and in line with simplification trends observed at the lexical level. However, using Dependency Locality Theory (DLT) to assess syntactic complexity, no general simplification effect via shorter dependency length was found, but rather a syntactic source speech influence shining-through. Average dependency length for all dependency relations and the percentage of very short dependency relations showed no trend of simplification in interpreting. Only the average maximum dependency length was shorter in interpreting than comparable originals, which can be explained by shorter sentences in interpreting generally. The only slight simplification trend using DLT was found in the qualitative analysis, showing that long interpreted sentences in German tend to use shorter dependency relations than the German originals.

Source language traces that were found in the comparable dependency length analysis, were also evident in the parallel analyses at the lexical and syntactic levels. In a parallel approach focussing on simplification from the source to the target speeches, it became evident that interpretese is characterised not only by simplification patterns, but also by shining-through of source language properties at multiple levels. Furthermore, the parallel perspective shed light on the context in which simplification occurs. As previous research shows processing and retrieval differences for content and function words (Seifart et al., 2018), investigating nouns and verbs as the main carriers of the message was the focus of this analysis. Simplification trends were noted for the low and high range of surprisal (Sections 5.2.3 and 5.2.4). Low surprisal verbs and English low surprisal nouns in interpreting were found to be triggered by low surprisal input, added as a low surprisal filler or a low surprisal pragmatic discourse marker, and particularly by using standardised translation solutions for a diverse range of input. While highly predictable words are produced faster, linked to ease in lexical access and high pre-activation (Bell et al., 2003; Malisz et al., 2018; Huettig et al., 2022), low surprisal words are also associated with lower cognitive effort on the comprehension side (Demberg and Keller, 2008; Smith and Levy, 2013), i.e. the interpreting product becomes more simplified when a higher proportion of low surprisal words are used.

On the other hand, unexpected input such as high surprisal lexical words from the source speech requires high processing costs for the interpreter (Demberg and Keller, 2008; Smith and Levy, 2013), who is not able to resort to standard material,

but rather has to come up with an equally unexpected (high surprisal) translation solution in the target. Compared to low surprisal words, high surprisal words are at a planning disadvantage as the previous context does not make them as predictable (Pollkläsener et al., 2025). They are linked to less pre-activation and more difficulty in planning. As a means of dealing with these high processing costs, the use of cognates was observed, in line with the cognate facilitation effect as described in the literature (Shlesinger and Malkiel, 2005; Shlesinger and Ordan, 2012; Defrancq, 2015; Chmiel et al., 2021, 2023). In such cases, interpreters choose a translation equivalent that is lexically very close to the source speech input, however, leads to unexpected lexical items in interpreting with high surprisal. This trend is rather the opposite of simplification and can be clearly linked to source speech influence. Furthermore, the study showed that when interpreters struggle with unexpected input, such as high surprisal lexical words, they also tend to simplify the target speech by omitting them in the interpreting product.

While surprisal values as obtained here do not encode syntactic structure, syntactic structure certainly also influences surprisal (Slaats and Martin, 2025). As lexis and syntax are clearly linked, it is not surprising that the parallel analysis showed similar tendencies at the lexical and syntactic level. Traces of the source speeches at the macro-level were found in the alignment analysis (Section 6.2.2) with interpreters generally producing one target sentence for one source segment. This observation of maintaining the general macrostructure of the source speech was previously stated in the literature (Liontou, 2011). Moreover, the alignment analysis showed a clear trend of syntactic simplification by interpreters to split long source sentences into two or up to seven target sentences. Jiang and Jiang (2020) showed that long dependency distance input, as may occur in long source sentences, can create processing difficulties for the interpreters, who might resort to syntactic simplification to cope with the increased cognitive burden (Xu and Liu, 2023). A trend of syntactic simplification was also found in the qualitative DLM analysis. By producing several shorter complete sentences, source speech elements do not need to be stored in memory as long as compared to reproducing a long and complex source speech sentence. Breaking up long source sentences into several shorter ones was also mentioned previously in the literature by linking word level results to such an observation (He et al., 2016; Gast and Borges, 2023). However, this systematic analysis of manual source-to-target alignment of the entire EPIC-UdS corpus is the first study of this kind to our knowledge that shows such a result in a quantitative analysis, systematically comparing source items to target equivalents. Furthermore, it was quantitatively shown that these split target sentences sometimes use the same or similar sentence structure in subsequent sentences, which contribute to higher standardisation and simplification as defined here. While merging several source segments into one target segment was found to be rare, it should still be mentioned that from a syntactic point of view, a summary translation (Al-Khanji et al., 2000; Li, 2015) is not necessarily more complex.

A parallel perspective on dependency relations also revealed some syntactic simplification trends, however, only for some individual relations (Section 6.2.3). The analysis revealed that interpreters optimise their cognitive resources by producing shorter sentences, splitting the source input, and maintaining it in a fairly linear way. This is a more prominent pattern than minimising dependency length.

The lexical and syntactic studies in this thesis also revealed some clear language differences. Lexical simplification was found to be more prominent in English interpreting than in German interpreting. Average dependency length for almost all individual dependency relations in German interpreting was lower than for German originals. This aligns with a trend resulting from the alignment analysis, where interpreters more frequently split English source input into serval German target sentences compared to the reverse interpreting direction (German originals interpreted into English). German's freer word order may be one explanation, allowing for more flexibility in the sentence structure. Original English, on the other hand, already uses a more minimal dependency length structure than German (Gildea and Temperley, 2010; Futrell et al., 2015; Dyer, 2023), so the pressure to optimise dependency distances in cognitively challenging conditions such as simultaneous interpreting may not be as strong in the interpreting direction German into English.

The focus of this thesis was to find out whether the cognitively challenging production conditions in simultaneous interpreting would force interpreters to simplify the interpreting product. While this was generally confirmed, the studies carried out in this thesis also showed that interpreting strategies, theoretically discussed in the literature, are applied in the interpreting process and are observable in the interpreting product. Interpreting syntactic symmetry (Seeber, 2011) is the preferred interpreting strategy at the sentence level. The strategy of selection and deletion or omission (Kalina, 1998; Gile, 1995; Setton, 1999; Korpal, 2012; Jones, 1998; Al-Khanji et al., 2000) became evident at various levels. It was mainly observed for high surprisal verbs and nouns, but also qualitatively in the alignment analysis and parallel dependency relations analysis. Summary translations (Kalina, 1998; Li, 2015) are rarely used when several source sentences are merged into one target sentence.

To summarise the main findings, it can be stated that, generally, interpreters tend to encode their output rational to cope with restricted time and cognitive resources. Optimal encoding in interpreting is linked to simplification as well as maintaining certain source texts characteristics. Simplification of the interpreting product was found at the lexical level with traditional measures, the information-theoretic measure of surprisal, as well as syntactically via shorter sentences, and splitting of source input. However, no general trend of Dependency Length Minimisation (DLM) (Gibson, 2000) was found, but only for some individual dependency length minimisation measures. Simplification trends are sometimes counterweighted by shining-through

of source language properties with the use of cognates at the lexical level or maintaining source speech structure. These strategies seem to be cognitively more efficient under certain conditions than simplification mechanisms. Generally, it can be concluded that properties of interpretese are linked to production facilitation instead of audience design.

7.2 Conclusion and outlook

This thesis aimed to investigate simultaneous interpreting as an understudied language variety. The main contributions can be summarised as follows. With **EPIC-UdS** (Przybyl et al., 2022b), this thesis provided **a valuable interpreting corpus resource** for an understudied language combination in corpus-based interpreting studies. With a focus on the language pair German and English (as source and target languages), the thesis gave **a better understanding of particular characteristics of simultaneous interpreting** with the finding of *interpretese* being particularly characterised by spoken language elements. This is in line with previous studies that include other language combinations (Shlesinger and Ordan, 2012; Karakanta et al., 2021; Gast and Borges, 2023). Previous findings on interpretese using traditional measures were mostly confirmed (Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2015; Ferraresi et al., 2018; Dayter, 2018; Kajzer-Wietrzny and Ivaska, 2020), while this thesis provided **more fine-grained measures using annotation on delivery mode and rate**. More importantly, this thesis used measures that were previously not, or only infrequently used in interpreting research. **Relative entropy** showed distinctive features of interpreted language in a data-driven approach. Frameworks that connect **predictability and memory** were applied to interpreting data, providing **novel insights into traces of the interpreting process** that are visible in interpreting product data. The information-theoretic measure of **surprisal as an indicator of the effort required to comprehend or produce language**, providing a direct link to cognition (Hale, 2001; Teich et al., 2020), was used to approximated lexical simplification. **Dependency Locality Theory was used to approximate syntactic processing** (Gibson, 2000; Liu, 2008). Furthermore, an analysis of **manual source-to-target sentence alignment of the entire EPIC-UdS corpus** showed that interpreters mostly maintain source speech input linearly, or split source input and simplify syntactically. Besides simplification, cognitive facilitation is achieved by **staying close to the source input**, syntactically as well as lexically. Generally, the main assumption that the cognitively challenging production condition in interpreting would require interpreters to encode their output optimally and this would leave traces of **lexical and syntactic simplification** in the interpreting product was largely confirmed, although more so in the English subcorpora than in German.

While this corpus-based approach looked at general traces of simplification at var-

ious levels, several limitations should also be mentioned. The approaches taken in this thesis studied general effects of the challenging production condition in simultaneous interpreting, however, it did not particularly focus on problem triggers or even translation errors. Omissions were observed in the qualitative studies, but it was not investigated whether such omissions would lead to loss of information or a lack of cohesion, which would lead to a less simplified product where the audience would be required to close the gap. In particular, the lexical studies showed that interesting phenomena occur at the lower and higher end of the surprisal range for lexical words, with additions of highly expected input, and omissions or shining-through of source language traces for unexpected input. These ends of the range for surprisal might distort the general result of simplification for the entire corpus. Specifically when expected output is added this may lead to more simplification, or less simplification when high surprisal input. Investigating the corpus without these low and high range of surprisal tokens, and investigating those separately, might give an even clearer result of contexts in which simplification occurs.

One limitation of this thesis is the way surprisal was modelled in the EPIC-UdS corpus variant used in this thesis. The language models used for calculating surprisal were built on n-gram models only based on EPIC-UdS corpus data (excluding the speech on which the model was applied). This approach was taken to reflect the particular setting and specificities of this spoken language corpus. However, it can be argued that the corpus at hand with 110 000 to 180 000 tokens per language is too small, and also slightly unbalanced with a larger share of interpreted speeches than originals in the dataset, to incorporate the entire linguistic and context knowledge of European Parliament speakers. To overcome this issue, Kunilovskaya et al. (2023b) use a strictly balanced dataset and use lemma surprisal to mitigate effects of German's rich morphology. Furthermore, surprisal can also be retrieved based on computational language models from larger corpora, using LLMs or Transformers such as Llama or GPT (Pollkläsener et al., 2025). Besides capturing the probability of a word based on the proceeding words, transformer surprisal may also capture "the influence of contextual constraints earlier in the sentence" (Pollkläsener et al., 2025) and may better reflect syntactic decisions made by the interpreters via the probability of the upcoming words based on a larger data basis. Nevertheless, having stated these downsides of the approach taken in this thesis, it can be said that the results of the studies in this thesis align with studies that use different modelling approaches for the same dataset (Kunilovskaya et al., 2023b; Pollkläsener et al., 2025), confirming the validity of the results obtained in this thesis.

Another limitation has to be stated for the decision taken on dependency parsing within this thesis. Filled pauses were included in the corpus data for parsing to stay as true as possible to the spoken nature of the data. However, the qualitative analysis in Chapter 6 showed parsing inaccuracies for certain dependency relations linked to filled pauses. As filled pauses also have a different status in a sentence and are not required to be maintained in memory until the dependency relation is closed, they

could be ignored for dependency parsing to give a more precise picture of dependency minimisation trends in interpreting.

This thesis aimed to combine a lexical and syntactic approach to cognition in simultaneous interpreting that captures prediction and memory aspects of the interpreting process. The methods were applied separately to the individual subcorpora, i.e. separately to source and target speeches, using monolingual language models. Furthermore, alignment was available only at the sentence level. Several of these downsides from the approaches of these thesis are overcome in the most recent release of the EPIC-UdS corpus (Kunilovskaya and Pollklaesener, 2026). The authors use monolingual GPT-2 models as well as bilingual machine translation models for estimating surprisal, remove filled pauses before parsing and reinsert them afterwards, and provide word-level alignment. While surprisal from GPT-2 models correlates well with experimental measures of processing difficulties in a monolingual setting (Oh and Schuler, 2023), few studies report on correlations of cross-lingual MT model surprisal with contradicting results (Lim et al., 2023; Kunilovskaya and Pollklaesener, 2026).

To fully reflect the nature of the interpreting process, with the source speech being heard and shortly afterwards the target speech being produced by the interpreter, time alignment on word level would be required. With this additional annotation, it would also be possible to better model memory processes and vary the number of tokens to be held in memory, depending on the actual EVS. Instead of modelling surprisal using a fixed context of 3 tokens (4-gram model), the context could be selected to reflect actual EVS. EVS in simultaneous interpreting is usually between one to three seconds, which corresponds to a larger context than 3 tokens, but larger time lags are also possible (Defrancq, 2015). Modelling the bilingual data with variable memory, using variable context, would come closer to the actual interpreting process and could give more detailed insights into the specific linguistic choices made by interpreters.

Having stated these limitations and possible future work, it can be concluded that this thesis provided a profound understanding of *interpretese* by applying novel methods in the field of corpus-based interpreting studies. It provides a solid description of general characteristics of interpreted language compared to original spoken language and specifically showed effects of production and memory optimisation, traceable as simplification in simultaneous interpreting.

Zusammenfassung

Simultandolmetschen ist eine Form der Sprachproduktion, die durch die besonderen Zwänge des Dolmetschprozesses geprägt wird. Diese Arbeit vermittelt ein fundiertes Verständnis der spezifischen sprachlichen Eigenschaften der Dolmetschsprache. Diese wird auch als „Interpretese“ bezeichnet. Simultandolmetschen ist eine kognitiv sehr anspruchsvolle Tätigkeit, zu der die gleichzeitige Koordination von Dekodierung der Ausgangssprache, Kodierung und Monitoring des zielsprachlichen Outputs sowie die zwischenzeitliche Speicherung von Ausgangs- und Zielelementen im Gedächtnis bis zum Abschluss der Produktion gehört (Gile, 1995). Diese Dissertation untersucht die These, dass Dolmetscher aufgrund der insgesamt kognitiv anspruchsvollen Aufgabe ihren Output optimal kodieren. Informationstheorie und rationale Kommunikation (Shannon, 1948) bilden den größeren Rahmen für die Untersuchungen in dieser Arbeit. Es wird angenommen, dass Sprecher den Informationsgehalt ihrer Äußerungen durch den Einsatz von entsprechenden linguistischen Mitteln so anpassen, dass die Botschaft ihrer Äußerung übertragen wird und gleichzeitig die kognitiven Anforderungen optimiert sind (Crocker et al., 2016). Übertragen auf den Dolmetschprozess wird erwartet, dass Dolmetscher, ebenso wie Sprecher in anderen Sprechsituationen, ihre Äußerung optimal kodieren. Aufgrund der besonders anspruchsvollen Produktionsbedingungen wird jedoch erwartet, dass Dolmetscher lexikalische und syntaktische Entscheidungen treffen, die sich von optimaler Kodierung in anderen Sprechsituationen unterscheidet. Um die Kapazität des Kanals nicht zu überschreiten wird erwartet, dass optimale Kodierung in der Verdolmetschung durch Vereinfachung (d.h. Simplifizierung) des Dolmetscherprodukts geprägt ist.

In der Dolmetschforschung werden kognitive Prozesse meist theoretisch diskutiert (Gerver, 1976; Moser, 1978; Setton, 1999; Gile, 2009, 2017; Camayd-Freixas, 2011; Seeber, 2011, 2013) oder experimentell untersucht (Seeber and Kerzel, 2012a; Ivanova, 2000; Bartłomiejczyk, 2007; Gumul, 2020). Diese Arbeit verfolgt einen korpusbasierten Ansatz um nicht nur punktuelle, prozessbedingte Effekte zu untersuchen, sondern vielmehr Simplifizierung im Dolmetschprodukt insgesamt zu analysieren. Grundlage für die Studien dieser Arbeit ist EPIC-UdS (Przybyl et al., 2022b), ein qualitativ hochwertiges Dolmetschkorpus basierend auf Reden, die in den Plenarsitzungen des Europäischen Parlaments gehalten wurden. Die englischen und deutschen Subkorpora von EPIC-UdS mit jeweils Originalreden und Verdolmetschung sind die Grundlage für die vergleichbaren und parallelen Studien in dieser Arbeit. Ein vergleichbarer Ansatz (Vergleich von originalsprachlichen und verdolmetschten Reden in der selben Sprache) und ein paralleler Ansatz (Vergleich von Originalreden und Verdolmetschungen dieser Originale) werden verfolgt, um (a) Simplifizierung in Verdolmetschungen im Vergleich zu den Normen der Zielsprache und (b) Simplifizierung in Verdolmetschungen im Vergleich zum ausgangssprachlichen Input zu untersuchen.

Diese Arbeit verwendet Methoden, die bisher in der Dolmetschforschung nicht, oder nur selten angewendet wurden. Dabei wird berücksichtigt, dass bei der Sprachverarbeitung, insbesondere unter schwierigen Produktionsbedingungen wie beim Simultandolmetschen, Erwartbarkeit und Gedächtnis (Optimierung) eine wichtige Rol-

le spielen. Zunächst verwendet die Arbeit das informationstheoretische Maß *Relative Entropie (Kullback-Leibler-Divergenz, KLD)* (Kullback and Leibler, 1951), um charakteristische Merkmale des Simultandolmetschens im Vergleich zur gesprochensprachlichen Originalen zu erkennen, und liefert eine datengestützte Beschreibung von verdolmetschter Sprache, ohne zuvor die zu untersuchenden Merkmale auszuwählen.

Mit dem Fokus auf Erwartbarkeit und Gedächtnis verwendet diese Arbeit Methoden zur Untersuchung von Simplifizierung, die einen direkten Bezug zu Kognition herstellen. Dies gilt insbesondere für den zweiten in dieser Arbeit verwendeten Rahmen. Das informationstheoretische Maß *Surprisal*, also der in einer sprachlichen Einheit kodierte Informationsgehalt (Hale, 2001), wird verwendet, um Erwartbarkeit der verwendeten lexikalischen Wörter zu analysieren. Surprisal ist hoch, wenn ein Wort in seinem Kontext eine geringe Wahrscheinlichkeit hat. Wörter mit niedrigem Surprisal gelten als sehr erwartbar im entsprechenden Kontext. Surprisal kann als Indikator für den Aufwand angesehen werden, der zum Verständnis oder zur Produktion von Sprache erforderlich ist (Levy, 2008; Smith and Levy, 2013; Bicknell and Levy, 2010; Demberg and Keller, 2008). Eher erwartbare Wörter, d.h. Wörter mit niedrigem Surprisal, werden hier mit einem höheren Grad an Simplifizierung in Verbindung gebracht, während weniger erwartbare Wörter weniger Simplifizierung bedeuten. Dieser informationstheoretische Ansatz wird durch traditionelle, korpusbasierte Ansätze zur Untersuchung von lexikalischer Simplifizierung, wie lexikalische Dichte, Type-Token-Ratio und lexikalische Variation, ergänzt (Laviosa, 1998; Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2012, 2015; Ferraresi et al., 2018; Dayter, 2018).

Der dritte in dieser Arbeit verwendete Ansatz ist Gedächtnisoptimierung über *Dependency Locality Theory (DLT)* (Gibson, 1998, 2000). Syntaktische Simplifizierung wird als linearer Abstand zwischen syntaktisch abhängigen Wörtern untersucht. Größere Abstände erfordern einen höheren kognitiven Aufwand, um sprachliche Elemente im Arbeitsgedächtnis zu speichern (Gibson, 2000; Liu, 2008). Kürzere Abstände, oder größere Lokalität, zwischen syntaktischen Elementen bedeuten geringere syntaktische Komplexität. Eine Minimierung dieser Abstandslänge *(Dependency Length Minimisation, DLM)* kann allgemein bei natürlichen Sprachen beobachtet werden (Gildea and Temperley, 2010; Temperley, 2007). Fürs Simultandolmetschen wird angenommen, dass ein besonderer Druck zur Gedächtnisoptimierung, und somit möglichst kurzer Zwischenspeicherung syntaktischer Elemente, besteht. Dieser syntaktische Ansatz über DLT wird mit einer Analyse von quell- und zielsprachlichen Satzalignierungen kombiniert, um zu untersuchen, ob und, wenn ja, wie Dolmetscher ausgangssprachliche Sätze in der Verdolmetschung splitten, zusammenfügen, oder ob die ausgangssprachliche Makrostruktur vom Dolmetscher beibehalten wird.

Die Arbeit ist wie folgt aufgebaut. Nach der Einleitung in Kapitel 1 wird in Kapitel 2 der theoretische Hintergrund eingeführt. Hier wird zunächst auf Simplifizierung im Kontext von schriftsprachlicher Translationsforschung eingegangen. Des Weiteren wird der Prozess des Simultandolmetschens, theoretische Dolmetschmodelle und -strategien, sowie Ansätze zur Beschreibung kognitiver Belastung beim Simultan-

dolmetschen vorgestellt. Anschließend werden Informationstheorie und Dependency Locality Theory (DLT) eingeführt, und deren bisher begrenzte Anwendung in der Dolmetschforschung aufgezeigt. Ein wichtiger Teil der theoretischen Grundlagen stellt das Teilkapitel zu *Interpretese*, also den bisher verfügbaren Erkenntnissen zu sprachlichen Besonderheiten von verdolmetschter Sprache, dar. Bisherige Studien zeigen ein sehr begrenztes Bild auf, das sich auf nur wenige Sprachkombinationen und ohne Deutsch als Ausgangs- und/oder Zielsprache stützt, und mit traditionellen korpusbasierten Methoden uneindeutige Ergebnisse liefert (Kajzer-Wietrzny, 2015; Sandrelli and Bendazzoli, 2005; Ferraresi et al., 2018; Dayter, 2018; Lv and Liang, 2019; Russo et al., 2012; Kajzer-Wietrzny and Ivaska, 2020; Xu and Li, 2022). Das Kapitel endet mit einer Zusammenfassung der theoretischen Grundlagen, sowie der Definition von Simplifizierung und der Forschungsfragen für die vorliegende Arbeit. Kapitel 3 führt EPIC-UdS (Przybyl et al., 2022b), das für diese Arbeit erstellte Dolmetschkorpus für die Sprachen Englisch und Deutsch mit den für diese Arbeit verwendeten Annotationen ein. Abschließend wird ein Überblick über die in dieser Arbeit angewandte Methodologie gegeben. Kapitel 4 bis 6 stellen mit den Studien den Hauptteil der Arbeit dar, wobei zu Beginn der Kapitel jeweils die einzelnen verwendeten Methoden im Detail erläutert, und anschließend die Untersuchungsergebnisse vorgestellt und diskutiert werden. Die Ergebnisse dieser Studien lassen sich wie folgt zusammenfassen (vgl. auch die Zusammenfassung in Kapitel 7):

Interpretese zeichnet sich in erster Linie durch Merkmale gesprochener Sprache aus, insbesondere durch gefüllte Pausen, Verstärkungspartikel, Deiktika, Diskursmarker und durch verbale, statt nominale Sprache. Das sind die wichtigsten Ergebnisse der explorativen Analyse dieser Arbeit (Kapitel 4), in der Interpretese mittels relativer Entropie (KLD) untersucht wird, und diese bestätigen die Ergebnisse anderer Autoren (Shlesinger and Ordan, 2012; Karakanta et al., 2021; Gast and Borges, 2023). Darüber hinaus werden diese Ergebnisse auch durch die Untersuchungen der qualitativen Analysen in dieser Arbeit bestätigt, in denen lexikalische und syntaktische Eigenschaften von Verdolmetschungen mittels Surprisal und DLT untersucht werden (Kapitel 5 und 6). Außerdem bestätigt die Analyse verschiedener Vortragsweisen der Originalreden (d.h. ob eine Rede im Parlament frei gehalten oder abgelesen wurde) in den lexikalischen Studien dieser Arbeit ebenfalls den Trend, dass Verdolmetschungen besonders von Mündlichkeit geprägt sind.

Der Schwerpunkt dieser Arbeit liegt auf der Untersuchung von Simplifizierung, welche auch bei schriftlichen Übersetzungen beobachtet wird (Baker, 1996; Blum and Levenston, 1978; Laviosa, 1998). Ausgehend von der Hypothese, dass der Dolmetschprozess den Dolmetscher in besondere Maße dazu zwingt, lexikalische und syntaktische Entscheidungen zu treffen, die für die Verarbeitung und das Gedächtnis von Vorteil sind und sich als Simplifizierung im Dolmetschprodukt nachweisen lassen, wird untersucht, ob Dolmetscher mehr kontextuell erwartbare Wörter verwenden, d.h. Wörter mit niedrigerem Surprisal als in Originalreden (Kapitel 5) und sich für syntaktische Komplexität gemessen über DLM-Indikatoren entscheiden (Kapitel 6).

Die Ergebnisse des vergleichbaren Ansatzes zeigen insgesamt Simplifizierungstendenzen im Dolmetschprodukt verglichen mit Originalreden in der selben Sprache, sowohl auf lexikalischer, als auch auf syntaktischer Ebene. Auf lexikalischer Ebene wird Simplifizierung zunächst unter Verwendung traditioneller Methoden beobachtet (Abschnitt 5.1). Lexikalische Simplifizierung zeigt sich insgesamt durch geringere lexikalische Dichte in der Verdolmetschung im Vergleich zu Originalreden in derselben Sprache, sowie bei spontanen Reden und bei Reden derselben Sprechgeschwindigkeit. Simplifizierung zeigt sich außerdem durch geringere lexikalische Variation in der Verdolmetschung, sowie durch kürzere Kodierungen auf Wort- und Nominalphrasenebene. Die Annäherung an Simplifizierung über Informationstheorie (Abschnitt 5.2) zeigt, dass Simplifizierung in der Verdolmetschung insgesamt durch die Verwendung von Inhaltswörtern mit niedrigerem Surprisal beobachtet werden kann. Dies zeigt sich insbesondere bei der Analyse aller Token, Verben (außer im Modus der spontanen Rede im Deutschen) und von Substantiven im Englischen (nicht jedoch im Deutschen). Darüber hinaus sind Verben mit niedrigem Surprisal generell häufiger in Verdolmetschungen als in Originalen beider Sprachen. Substantive mit niedrigem Surprisal treten jedoch häufiger in Originalen als in Verdolmetschungen auf. Verben und Substantive mit hohem Surprisal werden hingegen in beiden Sprachen häufiger in Originalreden als in Verdolmetschungen verwendet. Die Ergebnisse zeigen, dass Dolmetscher bei weniger überraschendem, erwartbarem Input eher auf standardisierte, erwartbare Übersetzungslösungen zurückgreifen. Bei unerwartetem (nominalem) Input, für die dem Dolmetscher kein häufig verwendetes Übersetzungsäquivalent zur Verfügung steht, lassen sich Dolmetscher auch von der Ausgangssprache beeinflussen. Die Verwendung von Kognaten kann die schnelle Produktion von Übersetzungslösungen unter Zeitdruck erleichtern (Chmiel et al., 2021).

Das Maß Surprisal spiegelt zumindest bis zu einem gewissen Grad auch lexikalische Variation im Kontext wider. Entsprechend stehen lexikalische Variation, gemessen mit traditionellen Methoden, und Surprisal lexikalischer Wörter als Erwartbarkeit im Kontext in Zusammenhang zueinander. Die Ergebnisse der Surprisalstudien ergänzen somit die Ergebnisse von geringerer lexikalischer Variation in Verdolmetschungen, die in den Studien mit traditionellen Methoden in dieser Arbeit festgestellt wurden, und stehen im Einklang mit den in der Literatur beschriebenen Ergebnissen (Russo et al., 2012; Kajzer-Wietrzny and Ivaska, 2020; Xu and Li, 2022).

Die Verwendung syntaktischer Methoden beim vergleichbaren Studiendesign führt zu leicht abweichenden Ergebnissen. Im Allgemeinen sind Sätze in der Verdolmetschung kürzer als in vergleichbaren Originalreden (Abschnitt 6.2.1), was mit den zuvor in der Literatur für andere Sprachen festgestellten Trends übereinstimmt (Bernardini et al., 2016; Liu et al., 2023) und mit den auf lexikalischer Ebene beobachteten Simplifizierungstendenzen im Einklang steht. Bei der Bewertung von syntaktischen Komplexität anhand Dependency Locality Theory (DLT) ist jedoch kein allgemeiner Simplifizierungseffekt durch syntaktische Elemente, die einen kürzeren Abstand zueinander haben (EN: shorter dependency length) festzustellen, sondern vielmehr

ein Einfluss der Ausgangssprache. Der durchschnittliche Abstand der abhängigen syntaktischen Elemente aller Relationen sowie der Prozentsatz sehr kurzer Abstände zeigen keinen Trend zur Vereinfachung beim Dolmetschen. Nur der durchschnittliche maximale Abstand abhängiger syntaktischer Elemente ist bei Verdolmetschungen kürzer als in vergleichbaren Originalen, was durch die allgemein kürzeren Sätze beim Dolmetschen erklärt werden kann. Der einzige leichte Simplifizierungstrend unter Verwendung von DLT wird in der qualitativen Analyse festgestellt, die zeigt, dass lange verdolmetschte Sätze im Deutschen tendenziell kürzere Relationen verwenden als dies in deutschen Originalreden der Fall ist.

Spuren der Ausgangssprache, die durch DLM-Indikatioren festgestellt wurden, zeigen sich auch in den parallelen Studien. In einem parallelen Ansatz, der Simplifizierung der Verdolmetschung im Bezug auf die dazugehörige Originalrede untersucht, wird deutlich, dass Verdolmetschungen nicht nur durch Simplifizierung gekennzeichnet sind, sondern auch durch das Durchscheinen von Eigenschaften der Ausgangsrede (Shining-through (Teich, 2003). Darüber hinaus beleuchtet die parallele Perspektive den Kontext, in dem Vereinfachungen auftreten.

Da frühere Studien Unterschiede bei der Verarbeitung und dem Abruf von Inhalts- und Funktionswörtern zeigen (Seifart et al., 2018), liegt der Schwerpunkt dieser Analysen auf der Untersuchung von Substantiven und Verben als Hauptträger einer Botschaft. Simplifizierungstendenzen sind sowohl für Inhaltswörter im niedrigem, als auch im hohen Surprisalbereich festzustellen (Abschnitte 5.2.3 und 5.2.4). Verben mit niedrigem Surprisal und englische Substantive mit niedrigem Surprisal in der Verdolmetschung werden sowohl durch Input mit niedrigem Surprisal ausgelöst, als Füllwörter mit niedrigem Surprisal oder pragmatische Diskursmarker mit niedrigem Surprisal hinzugefügt, oder auch insbesondere durch die Verwendung standardisierter Übersetzungslösungen für eine Vielzahl von ausgangssprachlichem Input genutzt. Während hochgradig erwartbare Wörter schneller produziert werden, was mit einem einfacherem lexikalischen Zugriff und einer hohen Voraktivierung zusammenhängt (Bell et al., 2003; Malisz et al., 2018; Huettig et al., 2022), werden Wörter mit niedrigem Surprisal auch mit geringerem kognitiven Aufwand auf der Verständnisseite in Verbindung gebracht (Demberg and Keller, 2008; Smith and Levy, 2013). Das Dolmetschprodukt gilt als vereinfacht, wenn ein höherer Anteil an Wörtern mit niedrigem Surprisal verwendet wird. Andererseits erfordert unerwarteter ausgangssprachlicher Input, wie z.B. Inhaltswörter mit hohem Surprisal, hohe Verarbeitungskosten für den Dolmetscher (Demberg and Keller, 2008; Smith and Levy, 2013), der nicht auf Standardmaterial zurückgreifen kann, sondern eine ebenso unerwartete Übersetzungslösung in der Zielsprache (mit hohem Surprisal) finden muss. Im Vergleich zu Wörtern mit niedrigem Surprisal sind Wörter mit hohem Surprisal in der Planung schwieriger, da sie aufgrund des vorherigen Kontexts weniger vorhersehbar sind (Pollkläsener et al., 2025). Sie werden mit einer geringeren Voraktivierung und größeren Schwierigkeiten bei der Planung in Verbindung gebracht. Als Mittel zum Umgang mit diesen hohen Verarbeitungskosten weisen die qualitativen Studien dieser Arbeit auf

einen kognitiven Erleichterungseffekt (EN: cognitive facilitation effet) (Chmiel et al., 2021) hin, bei dem Dolmetscher ein Übersetzungslösung wählen, die lexikalisch sehr nahe an der Ausgangssprache liegt. Die Verwendung von Kognaten in der Verdolmetschung wurde schon mehrfach in der Literatur beobachtet (Shlesinger and Malkiel, 2005; Shlesinger and Ordan, 2012; Defrancq, 2015; Chmiel et al., 2021, 2023) und führt allgemein zu unerwarteten lexikalischen Elementen im Dolmetschprodukt. Dies kann als das Gegenteil von Simplifizierung beschrieben werden, und wird eindeutig mit dem Einfluss der ausgangssprachlichen Rede in Verbindung gebracht. Darüber hinaus zeigt diese Arbeit, dass Dolmetscher, wenn sie mit unerwarteten ausgangssprachlichen Elementen wie Inhaltswörter mit hohem Surprisal konfrontiert werden, dazu neigen, die Zielsprache zu vereinfachen, indem sie diese in der Verdolmetschung weglassen.

Obwohl die hier verwendete Berechnung von Surprisalwerten die syntaktische Struktur nicht kodiert, beeinflusst die syntaktische Struktur sicherlich auch Surprisal (Slaats and Martin, 2025). Da Lexik und Syntax eindeutig miteinander verbunden sind, ist es nicht überraschend, dass die parallelen Analysen ähnliche Tendenzen für die lexikalischen und syntaktischen Studien aufzeigen. Spuren der Originalreden sind auch in der Analyse der Satzalignierung zu beobachten (Abschnitt 6.2.2). Es wird gezeigt, dass Dolmetscher in der Regel einen Zielsatz für ein ausgangssprachliches Segment produzieren. Eine Beibehaltung der allgemeinen Makrostruktur der Originalrede wurde zuvor auch in der Literatur festgestellt (Liontou, 2011). Darüber hinaus zeigt die Analyse der Satzalignierungen einen klaren Trend zu syntaktischer Vereinfachung, indem lange Ausgangssätze in der Verdolmetschung in zwei oder bis zu sieben Zielsätze gesplittet werden. Jiang and Jiang (2020) zeigen, dass lange syntaktische Abhängigkeiten, wie sie in langen Ausgangssätzen vorkommen können, zu Verarbeitungsschwierigkeiten für Dolmetscher führen können, die dann möglicherweise auf syntaktische Simplifizierung zurückgreifen, um diese erhöhte kognitive Belastung zu bewältigen (Xu and Liu, 2023). Ein Trend zur syntaktischen Simplifizierung wurde auch in der qualitativen DLM-Analyse festgestellt. Durch die Produktion mehrerer kürzerer, vollständiger Sätze müssen Elemente der Ausgangssprache nicht so lange im Gedächtnis gespeichert werden wie bei der Reproduktion eines langen und komplexen Satzes der Ausgangssprache. Die Aufsplittung langer Ausgangssätze in mehrere kürzere Sätze wurde bereits indirekt in der Literatur gezeigt (He et al., 2016; Gast and Borges, 2023). Die systematische Analyse der manuellen Alignierungen von Ausgangssätzen und Zielsätzen des EPIC-UdS-Korpus ist jedoch unseres Wissens nach die erste Studie dieser Art, die ein solches Ergebnis in einer quantitativen Analyse zeigt und Ausgangselemente systematisch mit Zieläquivalenten vergleicht. Darüber hinaus wird quantitativ gezeigt, dass diese gesplitteten, kürzeren Zielsätze auch die gleiche oder ähnliche Satzstruktur in nachfolgenden Sätzen verwenden, was zu einer höheren Standardisierung und Vereinfachung im hier definierten Sinne beiträgt. Obwohl die Zusammenführung mehrerer ausgangssprachlicher Segmente zu einem Zielsegment selten vorkommt, sollte dennoch erwähnt werden, dass eine zusammenfassende Über-

setzung (Al-Khanji et al., 2000; Li, 2015) aus syntaktischer Sicht nicht unbedingt komplexer sein muss.

Eine parallele Betrachtung der Abhängigkeitsrelationen (EN: dependency relations) zeigt ebenfalls einige Tendenzen zur syntaktischen Vereinfachung auf, jedoch nur für einzelne Relationen (Abschnitt 6.2.3). Daher kann festgestellt werden, dass DLM nicht die Hauptstrategie ist, mit der Dolmetscher ihre kognitiven Ressourcen optimieren. Die Verwendung kürzerer Sätze durch Aufspitten der Ausgangssätze und die relativ lineare Beibehaltung des ausgangssprachlichen Inputs ist die präferierte Strategie zur Ressourcenoptimierung.

Die lexikalischen und syntaktischen Untersuchungen in dieser Arbeit zeigen auch einige deutliche sprachspezifische Unterschiede auf. Es wird festgestellt, dass lexikalische Simplifizierung in englischen Verdolmetschungen stärker ausgeprägt ist als in deutschen Verdolmetschungen, und dass die durchschnittlichen Abstände für fast alle einzelnen Abhängigkeitsrelationen in der deutschen Verdolmetschung geringer ist als in deutschen Originalen. Dies steht im Einklang mit einem Trend, der aus der Analyse der Satzalignierungen hervorgeht. Es wurde gezeigt, das Dolmetscher englische Ausgangssätze häufiger in mehrere deutsche Zielsätze aufsplitten als in der umgekehrten Dolmetschrichtung (deutsche Originale, die ins Englische verdolmetscht werden). Eine mögliche Erklärung dafür ist die freiere Wortstellung im Deutschen, die mehr Flexibilität in der Satzstruktur zulässt. Englische Originale verwenden hingegen bereits optimiertere Strukturen mit geringeren Abständen von abhängigen syntaktischen Elementen als deutsche Originale (Gildea and Temperley, 2010; Futrell et al., 2015; Dyer, 2023), sodass der Druck, die Abstände unter kognitiv anspruchsvollen Bedingungen wie beim Simultandolmetschen zu optimieren, in der Dolmetschrichtung Deutsch ins Englische möglicherweise nicht so stark ist.

Zusammenfassend lässt sich sagen, dass Dolmetscher im Allgemeinen dazu neigen, ihren Output optimal zu kodieren, um mit begrenzten Zeit- und kognitiven Ressourcen zurechtzukommen. Simplifizierung in der Verdolmetschung wurde auf lexikalischer Ebene mit traditionellen Methoden, dem informationstheoretischen Maß Surprisal sowie syntaktisch durch kürzere Sätze und die Aufsplittung von Ausgangssätzen festgestellt. Es wurde jedoch kein allgemeiner Trend zur DLM (Gibson, 2000) festgestellt, sondern nur bei einzelnen Relationen eine Verringerung der Abstände abhängiger syntaktischer Elemente. Simplifizierungstendenzen werden teils durch das Durchscheinen von Eigenschaften der Ausgangssprache, wie durch die Verwendung von Kognaten auf lexikalischer Ebene oder die Beibehaltung der ausgangssprachliche Makrostruktur auf Satzebene, ausgeglichen. Diese Strategien scheinen unter bestimmten Bedingungen kognitiv effizienter zu sein als Simplifizierungsmechanismen. Generell lässt sich feststellen, dass Simplifizierung und Shinging-through als Eigenschaften von Verdolmetschungen eher sprecher- als hörerorienteiert sind.

Das Ziel dieser Arbeit war es, das Simultandolmetschen als eine wenig erforschte Sprachvariante zu untersuchen. Die wichtigsten Beiträge lassen sich wie folgt zusammenfassen: Mit **EPIC-UdS** (Przybyl et al., 2022b) lieferte diese Arbeit **ei-**

ne wertvolle Dolmetschkorpusressource für eine wenig erforschte Sprachkombination in der korpusbasierten Dolmetschforschung. Mit dem Schwerpunkt auf der Sprachkombination Deutsch und Englisch (als Ausgangs- und Zielsprache) ermöglichte die Arbeit ein besseres Verständnis der besonderen Merkmale von verdolmetschter Sprache, wobei festgestellt wurde, dass sich Dolmetschsprache insbesondere durch Elemente gesprochener Sprache auszeichnet. Dies steht im Einklang mit früheren Studien, die andere Sprachkombinationen untersuchen (Shlesinger and Ordan, 2012; Karakanta et al., 2021; Gast and Borges, 2023). Erkenntnisse früherer Studien zu Interpretese unter Verwendung traditioneller Messmethoden wurden größtenteils bestätigt (Sandrelli and Bendazzoli, 2005; Kajzer-Wietrzny, 2015; Ferraresi et al., 2018; Dayter, 2018; Kajzer-Wietrzny and Ivaska, 2020), während diese Arbeit **detailliertere Analysen unter Verwendung von Annotationen zu Vortragstyp und -geschwindigkeit** liefert. Vor allem aber verwendet diese Arbeit Ansätze, die zuvor in der Dolmetschforschung nicht oder nur selten angewandt wurden. Die Methode **relative Entropie** zeigte in einem datengesteuerten Ansatz charakteristische Merkmale der Dolmetschsprache. Methoden, die **Vorhersagbarkeit und Gedächtnis** miteinander verbinden, wurden auf Dolmetschdaten angewendet und liefern **neuartige Einblicke in Spuren des Dolmetschprozesses**, die im Dolmetschprodukt sichtbar sind. Das informationstheoretische Maß **Surprisal als Indikator für den Aufwand, der zum Verstehen oder Produzieren von Sprache erforderlich ist**, stellt einen direkten Bezug zur Kognition her (Hale, 2001; Teich et al., 2020) und wurde zur Annäherung an lexikalische Simplifizierung verwendet. Die **Dependency Locality Theory wurde verwendet, um syntaktische Simplifizierung zu approximieren** (Gibson, 2000; Liu, 2008). Darüber hinaus zeigte eine Analyse **manueller Satzalignierungen von Ausgangs- und Zielsätzen des gesamten EPIC-UdS-Korpus**, dass Dolmetscher die Makrostruktur der Ausgangsrede meist linear beibehalten oder Sätze der Ausgangsrede aufsplitten und syntaktisch vereinfachen. Neben Simplifizierung wird kognitive Erleichterung erreicht, indem Dolmetscher sowohl syntaktisch als auch lexikalisch **nahe an der Ausgangsrede bleiben**. Im Allgemeinen wurde die Hauptannahme, dass die kognitiv anspruchsvollen Produktionsbedingungen beim Simultandolmetschen von Dolmetschern eine optimale Kodierung ihrer Ausgabe erfordern und dies Spuren von **lexikalischer und syntaktischer Simplifizierung** im Dolmetschprodukt hinterlassen würde, weitgehend bestätigt. Des Weiteren wurden auch **Unterschiede im Sprachpaar** mit mehr Simplifizierungstendenzen im Englischen als im Deutschen beobachtet.

Bibliography

Ahrens, B. (2024). Cognitive models of interpreting. In Mellinger, C. D., editor, *The Routledge Handbook of Interpreting and Cognition*, pages 52–69. Routledge.

Al-Khanji, R., El-Shiyab, S., and Hussein, R. (2000). On the use of compensatory strategies in simultaneous interpretation. *Meta*, 45(3):548–557.

Alonso-Bacigalupe, L. (2010). Information processing during simultaneous interpretation: A three-tier approach. *Perspectives: Studies in Translatology*, 18:39–58.

Arnold, J., Losongco, A., Wasow, T., and Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76:28–55.

Aston, G. (2018). Acquiring the language of interpreters: A corpus-based approach. In *Making way in corpus-based interpreting studies*, volume 1 of *New Frontiers in Translation Studies*, pages 83–96. Springer.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In Baker, M., Francis, G., and Tognini-Bonelli, E., editors, *Text and Technology: In honour of John Sinclair*, pages 232–250. John Benjamins Publishing Company, Amsterdam.

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In Somers, H., editor, *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, page 175–186. John Benjamins, Amsterdam and Philadelphia.

Balling, L. W. and Kizach, J. (2017). Effects of surprisal and locality on danish sentence processing: An eye-tracking investigation. *Journal of psycholinguistic research*, 46:1119–1136.

Baroni, M. and Bernardini, S. (2006). A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text. *Literary and Linguistic Computing*, 21(3):259–274.

Barr, D. J. (2001). Trouble in mind: paralinguistic indices of effort and uncertainty in communication. *Oralité et gestualité: interactions et comportements multimodaux dans la communication: actes du colloque ORAGE 2001*.

Bartek, B., Lewis, R., Vasishth, S., and Smith, M. (2011). In Search of On-Line Locality Effects in Sentence Comprehension. *Journal of experimental psychology. Learning, memory, and cognition*, 37:1178–1198.

Bartłomiejczyk, M. (2006). Strategies of simultaneous interpreting and directionality. *Interpreting*, 8(2):149–174.

Bartłomiejczyk, M. (2007). Introspective methods in conference interpreting research. In *Challenging Tasks for Psycholinguistics in the New Century. Proceedings of the 7th Congress of International Society of Applied Psycholinguistics*. Katowice: Oficyna Wydawnicza Wacław Walasek.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in english conversation. *The Journal of the Acoustical Society of America*, 113(2):1001–1024.

Bendazzoli, C. and Sandrelli, A. (2005). An Approach to corpus-based interpreting studies: Developing EPIC (European Parliament Interpreting Corpus). In Gerzymisch-Arbogast, H. and Nauert, S., editors, *MuTra2005 - Challenges of Multidimensional Translation*. Proceedings of the Marie Curie Euroconferences. Saarbrücken. May 2005.

Bernardini, S., Ferraresi, A., and Miličević, M. (2016). From EPIC to EPTIC — exploring simplification in interpreting and translation. *Target*, 28/1:61–86.

Bernardini, S., Ferraresi, A., Russo, M., Collard, C., and Defrancq, B. (2018). Building interpreting and intermodal corpora: A how-to for a formidable task. In *Making way in corpus-based interpreting studies*, volume 1 of *New Frontiers in Translation Studies*, pages 21–42. Springer, Singapore.

Bicknell, K. and Levy, R. (2010). A rational model of eye movement control in reading. In *Annual Meeting of the Association for Computational Linguistics*.

Bizzoni, Y., Juzek, T. S., España-Bonet, C., Dutta Chowdhury, K., van Genabith, J., and Teich, E. (2020). How Human is Machine Translationese? Comparing Human and Machine Translations of Text and Speech. In *Proceedings of the 17th International Conference on Spoken Language Translation*, pages 280–290, Online. Association for Computational Linguistics.

Bizzoni, Y. and Lapshinova-Koltunski, E. (2021). Measuring Translationese across Levels of Expertise: Are Professionals more Surprising than Students? In *Proceedings of the 23rd NoDaLiDa*, pages 53–63, Online. ACL.

Bizzoni, Y., Przybyl, H., and Teich, E. (2021). Cutting semantic corners? Patterns of lexical simplification in interpreting vs. translation. In Castagnoli, S., Bernardini, S., Ferraresi, A., and Miličević Petrović, M., editors, *Using Corpora in Contrastive and Translation Studies Conference (6th Edition)*.

Bizzoni, Y. and Teich, E. (2019). Analyzing variation in translation through neural semantic spaces. In Serge Sharoff, P. Z. and Rapp, R., editors, *Proceedings of the 12th Workshop on Building and Using Comparable Corpora at Recent Advances in Natural Language Processing (RANLP) - Special topic: Neural Networks for Building and Using Comparable Corpora, Varna, Bulgaria*.

Blum, S. and Levenston, E. A. (1978). Universals of lexical simplification. *Language Learning*, 28(2):399–415.

Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*, pages 17–35. Gunter Narr, Tüübingen.

Brants, S., Dipper, S., Eisenberg, P., Hansen-Schirra, S., König, E., Lezius, W., Rohrer, C., Smith, G., and Uszkoreit, H. (2004). TIGER: Linguistic interpretation of a German corpus. *Research on Language and Computation*, 2:597–620.

Broersma, M., Carter, D., and Acheson, D. J. (2016). Cognate costs in bilingual speech production: Evidence from language switching. *Frontiers in Psychology*, 7.

Brysbaert, M., Mandera, P., and Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1):45–50.

Camayd-Freixas, E. (2011). Cognitive theory of simultaneous interpreting and training. In *Proceedings of the 52th Conference of the American Translators Association*, New York.

Carl, M. (2021). Information and entropy measures of rendered literal translation. In Carl, M., editor, *Explorations in Empirical Translation Process Research*, pages 113–140. Springer International Publishing, Cham.

Carl, M. and Schaeffer, M. (2017). Why Translation is Difficult: A Corpus-Based Study of Non-Literality in Post-Editing and From-Scratch Translation. *HERMES - Journal of Language and Communication in Business*, 2017(56):43–57.

Carl, M., Schaeffer, M., and Bangalore, S. (2016). The CRITT Translation Process Research Database. In Carl, M., Bangalore, S., and Schaeffer, M., editors, *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*, pages 13–54. Springer International Publishing, Cham.

Carter, D. and Inkpen, D. (2012). Searching for poor quality machine translated text: Learning the difference between human writing and machine translations. In Kosseim, L. and Inkpen, D., editors, *Advances in Artificial Intelligence – 25th Canadian Conference on Artificial Intelligence*, Lecture Notes in Computer Science, vol. 7310, pages 49–60. Springer.

Cecot, M. (2001). Pauses in simultaneous interpretation: a contrastive analysis of professional interpreters' performances. *Interpreters' Newsletter*, 11.

Chen, S. (2017). The construct of cognitive load in interpreting and its measurement. *Perspectives*, 25(4):640–657.

Chesterman, A. (2004). Beyond the particular. In Mauranen, A. and Kujamäki, P., editors, *Translation Universals: Do they exist?*, pages 33–49. John Benjamins, Amsterdam & Philadelphia.

Chesterman, A. (2017). *Reflections on Translation Theory: Selected papers 1993 - 2014*. John Benjamins.

Chmiel, A., Janikowski, P., Koržinek, D., Lijewska, A., Kajzer-Wietrzny, M., Jakubowski, D., and Plevoets, K. (2023). Lexical frequency modulates current cognitive load, but triggers no spillover effect in interpreting. *Perspectives*, 32(5):905–923.

Chmiel, A., Kajzer-Wietrzny, M., Korinek, D., Jakubowski, D., and Janikowski, P. (2024). Syntax, stress and cognitive load, or on syntactic processing in simultaneous interpreting. *Translation, Cognition & Behavior*, 7(1):22–47.

Chmiel, A., Kajzer-Wietrzny, M., Koržinek, D., and Janikowski, P. (2021). Cross-linguistic similarities in lexis: examining cognate activation through temporal and accuracy data from the Polish Interpreting Corpus (PINC). In Castagnoli, S., Bernardini, S., Ferraresi, A., and Miličević Petrović, M., editors, *Using Corpora in Contrastive and Translation Studies Conference (6th Edition)*, Bertinoro (Italy).

Chmiel, A., Koržinek, D., Kajzer-Wietrzny, M., Janikowski, P., Jakubowski, D., and Polakowska, D. (2022). Fluency parameters in the Polish Interpreting Corpus (PINC). In Kajzer-Wietrzny, M., Ferraresi, A., Ilmari, I., and Bernardini, S., editors, *Mediated discourse at the European Parliament: Empirical investigations*, chapter 3. Language Science Press.

Chmiel, A. and Lijewska, A. (2019). Syntactic processing in sight translation by professional and trainee interpreters: Professionals are more time-efficient while trainees view the source text less. *Target*, 31.

Chmiel, A., Szarkowska, A., Korzinek, D., Lijewska, A., Dutka, L., and Marasek, K. (2017). Ear–voice span and pauses in intra- and interlingual respeaking: An

exploratory study into temporal aspects of the respeaking process. *Applied Psycholinguistics*, 38(5):1201–1227.

Christodoulides, G. (2013). Prosodic features of simultaneous interpreting. In *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*, pages 33–37.

Christoffels, I. K., de Groot, A. M., and Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, 54(3):324–345.

Christoffels, I. K. and de Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In Kroll, J. F. and de Groot, A. M. B., editors, *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.

Clark, H. H. and Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1):73–111.

Collard, C., Przybyl, H., and Defrancq, B. (2019). Interpreting into an SOV language: memory and the position of the verb. A corpus-based comparative study of interpreted and non-mediated speech. *META*, 63(3):695–716.

Collins, M. (2014). Information density and dependency length as complementary cognitive models. *Journal of psycholinguistic research*, 43:651–681.

Conklin, K. and Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics*, 29(1):72–89.

Corpas Pastor, G. (2008). *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Peter Lang.

Corpas Pastor, G., Mitkov, R., Afzal, N., and Pekar, V. (2008). Translation universals: do they exist? A corpus-based NLP study of convergence and simplification. In *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas: Research Papers*, pages 75–81, Waikiki, USA. Association for Machine Translation in the Americas.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2):213–238.

Crocker, M. W., Demberg, V., and Teich, E. (2016). Information Density and Linguistic Encoding (IDeaL). *KI*, 30(1):77–81.

Crossley, S. A., Louwerse, M. M., McCarthy, P. M., and McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, 91(1):15–30.

Cuetos, F., Alvarez, B., González-Nosti, M., Méot, A., and Bonin, P. (2006). Determinants of lexical access in speech production: Role of word frequency and age of acquisition. *Memory & cognition*, 34:999–1010.

D. Alan Allport, B. A. and Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, 24(2):225–235. PMID: 5043119.

Dammalapati, S., Rajkumar, R., Ranjan, S., and Agarwal, S. (2021). Effects of duration, locality, and surprisal in speech disfluency prediction in English spontaneous speech. In Ettinger, A., Pavlick, E., and Prickett, B., editors, *Proceedings of the Society for Computation in Linguistics 2021*, pages 91–101, Online. Association for Computational Linguistics.

Dayter, D. (2018). Describing lexical patterns in simultaneous interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN). *FORUM: International Journal of Interpretation and Translation*.

Dayter, D. (2019). Collocations in non-interpreted and simultaneously interpreted english. In *New Empirical Perspectives on Translation and Interpreting*, pages 67–91. Routledge.

de Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., and Manning, C. D. (2014). Universal Stanford dependencies: A cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4585–4592, Reykjavik, Iceland. European Language Resources Association (ELRA).

Defrancq, B. (2015). Corpus-based research into the presumed effects of short EVS. *INTERPRETING*, 17(1):2:26–2:45.

Defrancq, B. (2016). Well, interpreters... a corpus-based study of a pragmatic particle used by simultaneous interpreters. In Corpas Pastor, G. and Seghiri Dominguez, M., editors, *Corpus-based approaches to translation and interpreting : from theory to applications*, volume 106 of *Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation*, pages 105–128. Peter Lang.

Defrancq, B. (2018). The European Parliament as a discourse community: Its role in comparable analyses of data drawn from parallel interpreting corpora. *The Interpreters' newsletter*, 23:115–132.

Defrancq, B. and Plevoets, K. (2018). Over-uh-load, filled pauses in compounds as a signal of cognitive load. In *Making way in corpus-based interpreting studies*, volume 1 of *New Frontiers in Translation Studies*, pages 43–64. Springer.

Defrancq, B., Plevoets, K., and Magnifico, C. (2015). Connective items in interpreting and translation: Where do they come from? In Romero-Trillo, J., editor, *Yearbook of corpus linguistics and pragmatics 2015: Current approaches to discourse and translation studies*, pages 195–222. Springer International Publishing, Cham.

Defrancq, Bart (2024). The dark load of simultaneous interpreting : interpreters doing it to themselves? In Mellinger, Christopher D., editor, *The Routledge handbook of interpreting and cognition*, pages 70–84. Routledge.

Degaetano-Ortlieb, S. and Teich, E. (2018). Using relative entropy for detection and analysis of periods of diachronic linguistic change. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 22–33, Santa Fe, New Mexico. Association for Computational Linguistics.

Degaetano-Ortlieb, S. and Teich, E. (2022). Toward an optimal code for communication: The case of scientific english. *Corpus Linguistics and Linguistic Theory*, 18(1):175–207.

Dell, G. S. and O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, 42(1):287–314.

Delogu, F., Crocker, M. W., and Drenhaus, H. (2017). Teasing apart coercion and surprisal: Evidence from eye-movements and erps. *Cognition*, 161:46–59.

Demberg, V. and Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.

Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, 43(3):449–470.

Dijkstra, T., val Hell, J. G., and Brenders, P. (2015). Sentence context effects in bilingual word recognition: Cognate status, sentence language, and semantic constraint. *Bilingualism: Language and Cognition*, 18(4):597–613.

Dillinger, M. (1994). Comprehension during interpreting: What do interpreters know that bilinguals don't? In Lambert, S. and Moser-Mercer, B., editors, *Bridging the Gap: Empirical research in simultaneous interpretation*, pages 155–190. John Benjamins.

Dimitrova, B. E. and Tiselius, E. (2009). Exploring retrospection as a research method for studying the translation process and the interpreting process. In Mees, I. M., Alves, F., and Göpferich, S., editors, *Methodology, technology and innovation in translation process research: a tribute to Arnt Lykke Jakobsen*, pages 109–134. Samfundslitteratur.

Doherty, M. (2006). *Structural Propensities: Translating nominal word groups from English into German*. John Benjamins.

Doi, K., Sudoh, K., and Nakamura, S. (2021). Large-scale English-Japanese simultaneous interpretation corpus: Construction and analyses with sentence-aligned data. In Federico, M., Waibel, A., Costa-jussà, M. R., Niehues, J., Stuker, S., and Salesky, E., editors, *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021)*, pages 226–235, Bangkok, Thailand (online). Association for Computational Linguistics.

Donato, V. (2003). Strategies adopted by student interpreters in SI: a comparison between the English-Italian and the German-Italian language-pairs. *The Interpreters' Newsletter*, 12:101–132.

Duff, A. (1981). *The third language: Recurrent problems of translation into English*. Pergamon, Oxford.

Dyer, A. T. (2023). Revisiting dependency length and intervener complexity minimisation on a parallel corpus in 35 languages. In Beinborn, L., Goswami, K., Muradoğlu, S., Sorokin, A., Kumar, R., Shcherbakov, A., Ponti, E. M., Cotterell, R., and Vylomova, E., editors, *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 110–119, Dubrovnik, Croatia. Association for Computational Linguistics.

Eisenberg, P. (1994). German. In König, E. and van der Auwera, J., editors, *The Germanic languages*, pages 349–387. Routledge, London.

Engelhardt, P. E., Barıs Demiral, S., and Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*, 77(2):304–314. Special Section: Aggression and peer victimization: Genetic, neurophysiological, and neuroendocrine considerations.

Ericsson, K. A. and Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 83(3):215–251.

Eskola, S. (2004a). Untypical Frequencies in Translated Language: A Corpus-Based Study on a Literary Corpus of Translated and Non-Translated Finnish. In Mauranen, A. and Kujamäki, P., editors, *Translation Universals: Do they exist?*, number 48 in Benjamins Translation Library, pages 83–99. John Benjamins Publishing Company.

Eskola, S. (2004b). Untypical patterns in translations. Issues on corpus methodology and synonymity. In Mauranen, A. and Kujamäki, P., editors, *Translation Universals: Do they exist?*, number 48 in Benjamins Translation Library, pages 101–126. John Benjamins Publishing Company.

Evert, S. and Neumann, S. (2017). The impact of translation direction on characteristics of translated texts. A multivariate analysis for English and German. In De Sutter, G., Lefer, M.-A., and Delaere, I., editors, *Empirical Translation Studies. New Theoretical and Methodological Traditions*, volume 300 of *Trends in Linguistics. Studies and Monographs (TiLSM)*, pages 47–80. Mouton de Gruyter, Berlin.

Fabricius-Hansen, C. (1996). Informational density: a problem for translation and translation theory. *Linguistics*, 34(3):521–566.

Fan, L. and Jiang, Y. (2019). Can dependency distance and direction be used to differentiate translational language from native language? *Lingua*, 224:51–59.

Fankhauser, P., Knappen, J., and Teich, E. (2014). Exploring and visualizing variation in language resources. In *Proceedings of the Language Resources and Evaluation Conference (LREC), May 2014, Reykjavik, Iceland*, pages 4125–4128.

Fellbaum, C. (2010). Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

Ferraresi, A., Bernardini, S., Petrović, M., and Lefer, M.-A. (2018). Simplified or not simplified? The different guises of mediated English at the European Parliament. *Meta*, 63(3):717–738.

Ferraresi, A. and Miličević, M. (2017). 5 Phraseological patterns in interpreting and translation. Similar or different? In Sutter, G. D., Lefer, M.-A., and Delaere, I., editors, *Empirical Translation Studies. New Methodological and Theoretical Traditions*, pages 157–182. De Gruyter Mouton, Berlin, Boston.

Ferreira, V. S. and Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4):296–340.

Ferrer-i-Cancho, R. (2006). Why do syntactic links not cross? *Europhysics Letters*, 76(6):1228–1234.

Frawley, W. (1984). Prolegomenon to a theory of translation. In Frawley, W., editor, *Translation: Literary, Linguistic and Philosophical Perspectives*, pages 159–175. University of Delaware Press, Newark.

Futrell, R. (2019). Information-theoretic locality properties of natural language. In Chen, X. and Ferrer-i Cancho, R., editors, *Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 2019)*, pages 2–15, Paris, France. Association for Computational Linguistics.

Futrell, R., Mahowald, K., and Gibson, E. (2015). Large-Scale Evidence of Dependency Length Minimization in 37 Languages. In *Proceedings of the National Academy of Sciences*, volume 112 (33), pages 10336–10341.

Gast, V. and Borges, R. (2023). Nouns, Verbs and Other Parts of Speech in Translation and Interpreting: Evidence from English Speeches Made in the European Parliament and Their German Translations and Interpretations. *Languages*, 8(1).

Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In Wollin, L. and Lindquist, H., editors, *Translation studies in Scandinavia: Poceedings from the Scandinavian Symposium on Translation Theory (SSOTT) II*, number 75 in Lund Studies in English, page 88–95. CWK Gleerup, Lund.

Genzel, D. and Charniak, E. (2002). Entropy rate constancy in text. In Isabelle, P., Charniak, E., and Lin, D., editors, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 199–206, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Gerritsen, M. and Stein, D., editors (1992). *Internal and External Factors in Syntactic Change*. De Gruyter Mouton, Berlin, Boston.

Gerver, D. (1976). Emprircal studies of simultaneous interpretation: A review and a model. In Brislin, R. W. and Anderson, R. B. W., editors, *Translation: Application and Research*. Gardner Press, New York.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., and Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5):389–407.

Gildea, D. and Temperley, D. (2007). Optimizing grammars for minimum dependency length. In Zaenen, A. and van den Bosch, A., editors, *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 184–191, Prague, Czech Republic. Association for Computational Linguistics.

Gildea, D. and Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, 34(2):286–310.

Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Benjamins translation library. Benjamins, Amsterdam [u.a.].

Gile, D. (1997). Conference interpreting as a cognitive management problem. *Applied Psychology-London-Sage-*, 3:196–214.

Gile, D. (1999). Testing the effort models' tightrope hypothesis in simultaneous interpreting-a contribution. *Hermes*, 23(1999):153–172.

Gile, D. (2004). Translation research versus interpreting research: Kinship, differences and prospects for partnership. In Schäffner, C., editor, *Translation Research and Interpreting Research: Traditions, Gaps and Synergies*, chapter 1, pages 10–34. Multilingual Matters, Bristol, Blue Ridge Summit.

Gile, D. (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *FORUM Revue internationale d'interprétation et de traduction / International Journal of Interpretation and Translation*, 6:59–77.

Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training: Revised edition*. John Benjamins.

Gile, D. (2017). Testing the Effort Models' Tightrope Hypothesis in Simultaneous Interpreting - A Contribution. *Hermes*, 23:153–172.

Goldman-Eisler, F. (1971). Segmentation of input in simultaneous translation. *Journal of Psycholinguistic Research*, 1(2):127–140.

Gordon, P. A. and Luper, H. L. (1989). Speech disfluencies in nonstutterers: Syntactic complexity and production task effects. *Journal of Fluency Disorders*, 14(6):429–445.

Granier, J. P., Robin, D. A., Shapiro, L. P., Peach, R. K., and Zimba, L. D. (2000). Measuring processing load during sentence comprehension: Visuomotor tracking. *Aphasiology*, 14(5-6):501–513.

Grice, H. P. (1975). *Logic and Conversation*, pages 41–58. Brill, Leiden, Niederlande.

Grodner, D. and Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, 29:261–90.

Gumul, E. (2006). Explicitation in simultaneous interpreting: A strategy or a by-product of language mediation? *Across Languages and Cultures*, 7(2):171–190.

Gumul, E. (2017). *Explicitation in Simultaneous Interpreting. A Study into Explicitating Behaviour of Trainee Interpreters*. Wydawnictwo Uniwersytetu Śląskiego.

Gumul, E. (2019). Evidence of cognitive effort in simultaneous interpreting: Process versus product data. *Beyond Philology An International Journal of Linguistics, Literary Studies and English Language Teaching*, 16(4):11–45.

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia*, 19:152–171.

Hahn, M., Degen, J., and Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. *Psychol Rev.*, 128(4):726–756.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 159–166, Stroudsburg, PA. Association for Computational Linguistics.

Halliday, M. (1989). *Spoken and Written Language*. Language education. Oxford University Press.

Halverson, S. (2003). The cognitive basis of translation universals. *Target*, 15:197–241.

Hansen-Schirra, S. (2011). Between normalization and shining-through. In Kranich, S., Becher, V., Höder, S., and House, J., editors, *Multilingual Discourse Production: Diachronic and Synchronic Perspectives*, pages 133–162. Hamburg Studies on Multilingualism 12.

Hawkins, J. (2014). *Cross-Linguistic Variation and Efficiency*. Oxford University Press.

Hawkins, J. A. (1995). *A Performance Theory of Order and Constituency*. Cambridge Studies in Linguistics. Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford University Press.

He, H., Boyd-Graber, J., and Daumé III, H. (2016). Interpretese vs. translationese: the uniqueness of human strategies in simultaneous interpretation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 971–976, San Diego, California. Association for Computational Linguistics.

Henderson, A., Goldman-eisler, F., and Skarbek, A. (1965). The common value of pausing time in spontaneous speech. *Quarterly Journal of Experimental Psychology*, 17(4):343–345.

Hild, A. (2011). Effects of linguistic complexity on expert processing during simultaneous interpreting. In Alvstad, C., Hild, A., and Tiselius, E., editors, *Methods and Strategies of Process Research. Integrative approaches in Translation Studies*, pages 249–267. John Benjamins Publishing Company.

Hoek, J., Zufferey, S., Evers-Vermeul, J., and Sanders, T. J. (2017). Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics*, 121:113–131.

House, J. (2008). Beyond intervention: Universals in translation? *trans-com*, 1(1):6–19.

House, J. (2016). Universals of translation? *Across Languages and Cultures*, 17:52–58.

Huettig, F., Audring, J., and Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, 224:105050.

Hughes, J. M., Foti, N. J., Krakauer, D. C., and Rockmore, D. N. (2012). Quantitative patterns of stylistic influence in the evolution of literature. *Proceedings of the National Academy of Sciences*, 109(20):7682–7686.

Hyönä, J., Tommola, J., and Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology Section A*, 48(3):598–612.

Ilisei, I., Inkpen, D., Corpas Pastor, G., and Mitkov, R. (2010). Identification of translationese: A machine learning approach. In *Lecture Notes in Computer Science book series, LNTCS*, volume 6008, pages 503–511.

Ivanova, A. (2000). The use of retrospection in research on simultaneous interpreting. In Trikkonen-Condit, S. and Jääskeläinen, R., editors, *Tapping and Mapping the Process of Translation and Interpreting*, volume 37, pages 27–52. Benjamins Translation Library.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

Jarvis, S. (2013). Defining and measuring lexical diversity. In Jarvis, S. and Daller, M., editors, *Vocabulary Knowledge: Human ratings and automated measures*, chapter 1, pages 13–44. John Benjamins.

Jiang, X. and Jiang, Y. (2020). Effect of dependency distance of source text on disfluencies in interpreting. *Lingua*, 243:102873.

Jing, Y., Blasi, D. E., and Bickel, B. (2022). Dependency length minimization and its limits: A possible role for a probabilistic version of the final-over-final condition. *Language*, 98(3):397–418.

Jones, R. (1998). *Conference interpreting explained.* St Jerome Publishing.

Kade, O. (1967). Zu einigen Besonderheiten des Simultandolmetschen. *Fachsprachen*, 11.1:8–17.

Kajzer-Wietrzny, M. (2012). *Interpreting universals and interpreting style.* PhD thesis, Adam Mickiewicz University, Poznan.

Kajzer-Wietrzny, M. (2015). Simplification in interpreting and translation. *Across Languages and Cultures*, 16(2):233–255.

Kajzer-Wietrzny, M. (2018). Interpretese vs. non-native language use: The case of optional that. In Russo, M., Bendazzoli, C., and Defrancq, B., editors, *Making Way in Corpus-based Interpreting Studies*, pages 97–113. Springer Singapore, Singapore.

Kajzer-Wietrzny, M. (2021). An intermodal approach to cohesion in constrained and unconstrained language. *Target-international Journal of Translation Studies*, 34:1:130–162.

Kajzer-Wietrzny, M. and Grabowski, L. (2021). Formulaicity in constrained communication: An intermodal approach. *MonTi Monografías de Traducción e Interpretación*, 13:148–183.

Kajzer-Wietrzny, M. and Ivaska, I. (2020). A multivariate approach to lexical diversity in constrained language1. *Across Languages and Cultures*, 21(2):169–194.

Kajzer-Wietrzny, M., Ivaska, I., and Ferraresi, A. (2024). Fluency in rendering numbers in simultaneous interpreting. *Interpreting*, 26(1):1–23.

Kalina, S. (1998). *Strategische Prozesse beim Dolmetschen: theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*. Language in performance. Narr, Tübingen.

Karakanta, A., Przybyl, H., and Teich, E. (2021). Exploring variation in translation with probabilistic language models. In *Corpora in translation research: recent advances and applications*, chapter 12, page 307–323. Benjamins Translation Library.

Karakanta, A., Vela, M., and Teich, E. (2018). Europarl-UdS: Preserving metadata from parliamentary debates. In Fišer, D., Eskevich, M., and de Jong, F., editors, *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018) Miyazaki, Japan, May 2018*, Paris, France. European Language Resources Association (ELRA).

Kathryn Bock, J. and Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21(1):47–67.

Kempen, G. and Harbusch, K. (2008). Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in german clauses. In Steube, A., editor, *The Discourse Potential of Underspecified Structures*, pages 179–192. De Gruyter, Berlin, New York.

Kim, K., Park, E.-J., Shin, J.-H., Kwon, O.-W., and Kim, Y.-K. (2017). Divergence-based fine pruning of phrase-based statistical translation model. *Computer Speech & Language*, 41:146–160.

Kirchhoff, H. (1976). Das Simultandolmetschen: Interdependenz der Variablen im Dolmetschprozess, Dolmetschmodelle und Strategien. In Drescher, H. W. und Scheffzek, S., editor, *Theorie und Praxis des Übersetzens und Dolmetschens*, pages 59–71. Peter Lang.

Klaudy, K. and Károly, K. (2005). Implicitation in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures*, 6:13–28.

Klingenstein, S., Hitchcock, T., and DeDeo, S. (2014). The civilizing process in london's old bailey. *Proceedings of the National Academy of Sciences*, 111(26):9419–9424.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the 10th Machine Translation Summit*, volume 5, pages 79–86, Phuket, Thailand. AAMT.

Kohn, K. and Kalina, S. (1996). The strategic dimension of interpreting. *Meta*, 41(1):118–138.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of psycholinguistic research*, 29:627–645.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, 66(1):143–162.

Korpal, P. (2012). Omission in simultaneous interpreting as a deliberate act. In Pym, A. and Orrego-Carmona, D., editors, *Translation Research Projects 4*, pages 103–111. Tarragona: Intercultural Studies Group.

Kravtchenko, E. and Demberg, V. (2015). Semantically underinformative utterances trigger pragmatic inferences. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., and Maglio, P. P., editors, *CogSci*. cognitivesciencesociety.org.

Kroll, J. F., Michael, E., Tokowicz, N., and Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, 18(2):137–171.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.

Kunilovskaya, M. and Lapshinova-Koltunski, E. (2020). Lexicogrammatic translationese across two targets and competence levels. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 4102–4112. European Language Resources Association.

Kunilovskaya, M., Mitkov, R., and Wandl-Vogt, E. (2023a). Cross-lingual Mediation: Readability Effects. In *Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2023)*, pages 33–43. INCOMA Ltd.

Kunilovskaya, M. and Pollklaesener, C. (2026). Epic-europarl-uds: Information-theoretic perspectives on translation and interpreting. In Piperidis, S., Bel, N., van den Heuvel, H., Ide, N., Krek, S., and Toral, A., editors, *Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026)*, pages 6998–7013, Palma, Mallorca, Spain. European Language Resources Association (ELRA).

Kunilovskaya, M., Przybyl, H., Lapshinova-Koltunski, E., and Teich, E. (2023b). Simultaneous interpreting as a noisy channel: How much information gets through. In Mitkov, R. and Angelova, G., editors, *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 608–618, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Kunz, K., Stoll, C., and Klüber, E. (2021). HeiCiC: A simultaneous interpreting corpus combining product and pre-process data. In Bizzoni, Y., Teich, E., España-Bonet, C., and van Genabith, J., editors, *Proceedings for the First Workshop on Modelling Translation: Translatology in the Digital Age*, pages 8–14. Association for Computational Linguistics.

Lapshinova-Koltunski, E. (2015). Variation in translation: Evidence from corpora. In Fantinuoli, C. and Zanettin, F., editors, *New directions in corpus-based translation studies.Translation and Multilingual Natural Language Processing (TMNLP)*, pages 79–99. Language Science Press.

Lapshinova-Koltunski, E., Pollkläsener, C., and Przybyl, H. (2022). Exploring Explicitation and Implicitation in Parallel Interpreting and Translation Corpora. *The Prague Bulletin of Mathematical Linguistics*, 119:5–22.

Lapshinova-Koltunski, E., Przybyl, H., and Bizzoni, Y. (2021). Tracing variation in discourse connectives in translation and interpreting through neural semantic spaces. In Braud, C., Hardmeier, C., Li, J. J., Louis, A., Strube, M., and Zeldes, A., editors, *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 134–142, Punta Cana, Dominican Republic and Online. Association for Computational Linguistics.

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, 43(4):557–570.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Levy, R. and Gibson, E. (2013). Surprisal, the pdc, and the primary locus of processing difficulty in relative clauses. *Frontiers in Psychology*, 4.

Levy, R. and Jaeger, T. F. (2006). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems*, 19:849–856.

Levy, R. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In Schölkopf, B., Platt, J., and Hoffman, T., editors, *Advances in Neural Information Processing Systems 19*. MIT Press, Cambridge, MA.

Levy, R. P. and Keller, F. (2013). Expectation and locality effects in german verb-final structures. *Journal of Memory and Language*, 68(2):199–222.

Li, X. (2015). Putting interpreting strategies in their place: Justifications for teaching strategies in interpreter training. *Babel*, 61(2):170–192.

Liang, J., Fang, Y., Lv, Q., and Liu, H. (2017). Dependency distance differences across interpreting types: Implications for cognitive demand. *Frontiers in Psychology*, 8.

Liang, Y. and Sang, Z. (2022). Syntactic and typological properties of translational language: A comparative description of dependency treebank of academic abstracts. *Lingua*, 273:103345.

Lijewska, A. and Chmiel, A. (2014). Cognate facilitation in sentence context – translation production by interpreting trainees and non-interpreting trilinguals. *International Journal of Multilingualism*, 12:1–18.

Lim, Z. W., Vylomova, E., Kemp, C., and Cohn, T. (2023). Predicting human translation difficulty with neural machine translation.

Liontou, K. (2011). Strategies in German-to-Greek Simultaneous Interpreting: A Corpus-Based Approach. *Gramma: Journal of Theory and Criticism*, 19.

Liu, H. (2008). Dependency Distance as a Metric of Language Comprehension Difficulty. *Journal of Cognitive Science*, 9:159–191.

Liu, H., Xu, C., and Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21:171–193.

Liu, K. and Afzaal, M. (2021). Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification. *PLOS ONE*, 16(6):1–20.

Liu, K., Liu, Z., and Lei, L. (2022). Simplification in translated chinese: An entropy-based approach. *Lingua*, 275:103364.

Liu, M., Schallert, D. L., and Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting*, 6(1):19–42.

Liu, Y., Cheung, A. K., and Liu, K. (2023). Syntactic complexity of interpreted, L2 and L1 speech: A constrained language perspective. *Lingua*, 286:103509.

Liu, Z. and Dou, J. (2023). Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: a cognitive perspective. *Frontiers in Psychology*, 14.

Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics*, 14:3–28.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15:474–496.

Lv, Q. and Liang, J. (2019). Is consecutive interpreting easier than simultaneous interpreting? – a corpus-based study of lexical simplification in interpretation. *Perspectives*, 27(1):91–106.

Ma, X., Hsu, Y.-Y., and Li, D. (2020). Exploring the impact of word order asymmetry on cognitive load during Chinese-English sight translation Evidence from eye-movement data. *Target*, 33.

Macháček, D., Žilinec, M., and Bojar, O. (2021). ESIC – Europarl Simultaneous Interpreting Corpus.

Malisz, Z., Brandt, E., Möbius, B., Oh, Y. M., and Andreeva, B. (2018). Dimensions of segmental variability: Interaction of prosody and surprisal in six languages. *Frontiers in Communication*, Volume 3.

Martínez, J. M. M. and Teich, E. (2017). Modeling routine in translation with entropy and surprisal: A comparison of learner and professional translations. In Cercel, L., Agnetta, M., and Lozano, M. T. A., editors, *Kreativität und Hermeneutik in der Translation*, pages 403–427. Narr Francke Attempto Verlag.

Mauranen, A. (2007). Universal Tendencies in Translation. In Anderman, G. and Rogers, M., editors, *Incorporating Corpora,The Linguist and the Translator*, chapter 3, pages 32–48. Multilingual Matters, Bristol, Blue Ridge Summit.

Mauranen, A. and Kujamäki, P., editors (2004). *Translation Universals: Do they exist?* John Benjamins.

Mazza, C. (2001). Numbers in simultaneous interpretation. *Linguistics*, pages 87–104.

Michaelov, J. A., Coulson, S., and Bergen, B. K. (2023). So cloze yet so far: N400 amplitude is better predicted by distributional information than human predictability judgements. *IEEE Transactions on Cognitive and Developmental Systems*, 15(3):1033–1042.

Moirón, B. V. and Tiedemann, J. (2006). Identifying idiomatic expressions using automatic word-alignment. In *Proceedings of the Workshop on Multi-word-expressions in a multilingual context.*

Monti, C., Bendazzoli, C., Sandrelli, A., and Russo, M. (2005). Studying directionality in simultaneous interpreting through an electronic corpus: EPIC (European Parliament Interpreting Corpus). *Meta*, 50(4).

Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In Gerver, D. and Sinaiko, H. W., editors, *Language Interpretation and Communication*, pages 353–368. Springer US, Boston, MA.

Neumann, S. (2003). *Textsorten und Übersetzen: eine Korpusanalyse englischer und deutscher Reiseführer*. Peter Lang Verlag, Berlin, Deutschland.

Nikolaev, D., Karidi, T., Kenneth, N., Mitnik, V., Saeboe, L., and Abend, O. (2020). Morphosyntactic predictability of translationese. *Linguistics Vanguard*, 6(1):1–12.

Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., and Zeman, D. (2016). Universal Dependencies v1: A multilingual treebank collection. In Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666, Portorož, Slovenia. European Language Resources Association (ELRA).

Oh, B.-D. and Schuler, W. (2023). Why Does Surprisal From Larger Transformer-Based Language Models Provide a Poorer Fit to Human Reading Times? *Transactions of the Association for Computational Linguistics*, 11:336–350.

Olohan, M. and Baker, M. (2000). Reporting that in translated English. Evidence for subconscious processes of explicitation. *Across Languages and Cultures*, 1:141–158.

Osborne, J. (2011). Fluency, complexity and informativeness in native and non-native speech. *International Journal of Corpus Linguistics*, 16(2):276–298.

Oster, K. (2017). The influence of self-monitoring on the translation of cognates. In Silvia Hansen-Schirra, O. C. . S. H., editor, *Empirical modelling of translation and interpreting*, pages 23–39. Language Science Press.

Pan, J. (2019). The Chinese/English political interpreting corpus (CEPIC): A new electronic resource for translators and interpreters. In *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*, pages 82–88, Varna, Bulgaria. Incoma Ltd., Shoumen, Bulgaria.

Park, Y. A. and Levy, R. (2009). Minimal-length linearizations for mildly context-sensitive dependency trees. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, page 335–343, USA. Association for Computational Linguistics.

Pickering, M. J. and Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, 3(4):136–141.

Pietrandrea, P., Kahane, S., Lacheret, A., and Sabio, F. (2014). The notion of sentence and other discourse units in spoken corpus annotation. In Mello, H. and Raso, T., editors, *Spoken corpora and Linguistic Studies*, pages 331–364. John Benjamins Publishing Company, Amsterdam-Philadelphia.

Pinochi, D. (2009). Simultaneous Interpretation of Numbers: Comparing German and English to Italian. An Experimental Study. *The Interpreters' Newsletter*, 2009(14):33–57.

Plevoets, K. and Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting: A corpus-based regression analysis. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, 11(2):202–224.

Plevoets, K. and Defrancq, B. (2018a). Lexis or parsing? A corpus-based study of syntactic complexity and its effect on disfluencies in interpreting. In *UCCTS 5 - Using Corpora in Contrastive and Translation Studies - September 2018*, Louvain-la-Neuve, Belgium.

Plevoets, K. and Defrancq, B. (2018b). The cognitive load of interpreters in the European Parliament: A corpus-based study of predictors for the disfluency uh(m). *Interpreting*, 20(1):1–28.

Plevoets, K. and Defrancq, B. (2020). Imported load in simultaneous interpreting: an assessment. In Muñoz Martín, R. and Halverson, S. L., editors, *Multilingual mediated communication and cognition*, The IATIS Yearbook, pages 18–43. Routledge.

Pöchhacker, F. (2004). *Introducing Interpreting Studies*. Routledge.

Pollkläsener, C., Kunilovskaya, M., and Teich, E. (2025). Surprisal explains the occurrence of filler particles in simultaneous interpreting. *SKASE Journal of Translation and Interpreting, Special issue: Empirical translation and interpreting studies*, 18(2):55–78.

Pollkläsener, C., Yung, F., and Lapshinova-Koltunski, E. (2024). Capturing variation of discourse relations in English parallel data through automatic annotation and alignment. *Across Languages and Cultures*, 25(2):288–309.

Przybyl, H., Karakanta, A., Menzel, K., and Teich, E. (2022a). Exploring linguistic variation in mediated discourse: translation vs. interpreting. In Kajzer-Wietrzny, M., Bernardinia, S., Ferraresi, A., and Ivaska, I., editors, *Empirical investigations into the forms of mediated discourse at the European Parliament*, Translation and Multilingual Natural Language Processing, page 191–218. Language Science Press, Berlin.

Przybyl, H., Lapshinova-Koltunski, E., Menzel, K., Fischer, S., and Teich, E. (2022b). EPIC UdS - creation and applications of a simultaneous interpreting corpus. In *Proceedings of the 13th International Language Resources and Evaluation Conference (LREC2022)*, pages 1193–1200, Marseille, France. European Language Resources Association.

Puurtinen, T. (2003). Genre-specific Features of Translationese? Linguistic Differences between Translated and Non-translated Finnish Children's Literature. *Literary and Linguistic Computing*, 18(4):389–406.

Pym, A. (2007). On shlesinger's proposed equalizing universal for interpreting. *Copenhagen studies in language*, pages 175–190.

Pym, A. (2009). On omission in simultaneous interpreting: Risk analysis of a hidden effort. *Efforts and Models in Interpreting and Translation Research*, pages 83–105.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

Rabadán, R., Labrador, B., and Ramón, N. (2009). Corpus-based contrastive analysis and translation universals: A tool for translation quality assessment. *Babel*, 55(4):303–328.

Rajkumar, R., van Schijndel, M., White, M., and Schuler, W. (2016). Investigating locality effects and surprisal in written English syntactic choice phenomena. *Cognition*, 155:204–232.

Reich, I. (2017). On the omission of articles and copulae in German newspaper headlines. *Linguistic variation*, 17:186–204.

Riccardi, A. (1998). Interpreting strategies and creativity. In Ann Beylard-Ozeroff, J. K. and Moser-Mercer, B., editors, *Selected Papers from the 9th International Conference on Translation and Interpreting, Prague, September 1995*, pages 171–180. Benjamins Translation Library.

Riccardi, A. (2005). On the evolution of interpreting strategies in simultaneous interpreting. *Meta*, 50(2):753–767.

Rubino, R., Lapshinova-Koltunski, E., and van Genabith, J. (2016). Information density and quality estimation features as translationese indicators for human translation classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 960–970, San Diego, California. Association for Computational Linguistics.

Ruiz, C., Paredes, N., Macizo, P., and Bajo, M. (2008). Activation of lexical and syntactic target language properties in translation. *Acta Psychologica*, 128(3):490–500. Bilingualism: Functional and neural perspectives.

Russo, M., Bendazzoli, C., Sandrelli, A., and Spinolo, N. (2012). The European Parliament Interpreting Corpus (EPIC): Implementation and developments. *Breaking ground in corpus-based Interpreting Studies*, 147:53–90.

Rutherford, A., Demberg, V., and Xue, N. (2017). A systematic study of neural discourse models for implicit discourse relation. In Lapata, M., Blunsom, P., and Koller, A., editors, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 281–291, Valencia, Spain. Association for Computational Linguistics.

Sandrelli, A. (2018). Interpreter-mediated football press conferences: A study on the questioning and answering strategies. In *Making way in corpus-based interpreting studies*, volume 1 of *New Frontiers in Translation Studies*, pages 185–204. Springer.

Sandrelli, A. and Bendazzoli, C. (2005). Lexical patterns in simultaneous interpreting: A preliminary investigation of EPIC (European Parliament Interpreting Corpus). In *Proceedings from the Corpus Linguistics Conference Series*.

Sandrelli, A., Bendazzoli, C., and Russo, M. (2010). European Parliament Interpreting Corpus (EPIC): Methodological issues and preliminary results on lexical patterns in simultaneous interpreting. *IJT-International Journal of Translation*, 22:165–203.

Santos, D. (1995). On grammatical translationese. In *Proceedings of the 10th Nordic Conference on Computational Linguistics*, pages 59–66. University of Helsinki.

Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W., and Carl, M. (2016). Word translation entropy: Evidence of early target language activation during reading for translation. In Carl, M., Bangalore, S., and Schaeffer, M., editors, *New Directions in Empirical Translation Process Research*, pages 183–210. Springer International Publishing, Cham.

Schaeffer, M., Nitzke, J., and Hansen-Schirra, S. (2019a). *Predictive Turn in Translation Studies: Review and Prospects*, pages 1–23. Springer Nature Switzerland AG.

Schaeffer, M., Nitzke, J., Tardel, A., Oster, K., Gutermuth, S., and Hansen-Schirra, S. (2019b). Eye-tracking revision processes of translation students and professional translators. *Perspectives*, 27(4):589–603.

Schäffner, C. and Adab, B. (2001a). The idea of the hybrid texts and translation: Contact as conflict. *Across Languages and Cultures*, 2(2):167–180.

Schäffner, C. and Adab, B. (2001b). The idea of the hybrid texts and translation revisited. *Across Languages and Cultures*, 2(2):277–302.

Schlesinger, M. (1991). Interpreter latitude vs. due process: Simultaneous and consecutive interpretation in multilingual trials. In Tirkkonen-Condit, S., editor, *Empirical Research in Translation and Intercultural Studies*, pages 147–155. Gunter Narr, Tübingen.

Schneider, R. and Lang, C. (2022). Das grammatische Informationssystem grammis – Inhalte, Anwendungen und Perspektiven. *Zeitschrift für germanistische Linguistik*, 50(2):407–427.

Schulz, E., Oh, Y. M., Andreeva, B., and Möbius, B. (2016). Impact of prosodic structure and information density on vowel space size. In *Proceedings of Speech Prosody*, pages 350–354, Boston, MA, USA.

Seeber, K. (2011). Cognitive load in simultaneous interpreting: Existing theories - new models. *Interpreting*, 13:176–204.

Seeber, K. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target*, 25(1):18–32.

Seeber, K. (2017). Interpreting at the european institutions: faster, higher, stronger. *CLINA : an interdisciplinary journal of ranslation, interpreting and intercultural communication*, 3(2):73–90.

Seeber, K. and Kerzel, D. (2012a). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, 16(2):228–242.

Seeber, K. G. and Kerzel, D. (2012b). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, 16(2):228–242.

Segalowitz, S. J. and Lane, K. C. (2000). Lexical access of function versus content words. *Brain and Language*, 75(3):376–389.

Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., Witzlack-Makarevich, A., de Jong, N. H., and Bickel, B. (2018). Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences*, 115(22):5720–5725.

Setton, R. (1999). *Simultaneous interpretation: A cognitive-pragmatic analysis*. John Benjamins, Amsterdam/Philadelphia.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Shao, Z. and Chai, M. (2020). The effect of cognitive load on simultaneous interpreting performance: an empirical study at the local level. *Perspectives*, 29(5):778–794.

Sheikh Al-Shabab, O. (1996). *Interpretation and the Language of Translation: Creativity and Convention in Translation.* Janus Pub., London, England, First English edition.

Shlesinger, M. (1989). *Simultaneous interpretation as a factor in effecting shifts in the position of texts in the oral-literate continuum.* Tel Aviv University. MA thesis, Tel Aviv.

Shlesinger, M. (1995). Shifts in cohesion in simultaneous interpreting. *The Translator*, 1(2):193–214.

Shlesinger, M. (1998). Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta*, 43:486–493.

Shlesinger, M. (2000). *Strategic Allocation of Working Memory and Other Attentional Resources in Simultaneous Interpreting.* Bar-Ilan University.

Shlesinger, M. (2009). Towards a definition of interpretese: An intermodal, corpus-based study. In Chesterman, A. and Gerzymisch-Arbogast, H., editors, *Efforts and Models in Interpreting and Translation Research: A tribute to Daniel Gile*, pages 237–253. Benjamins Translation Library.

Shlesinger, M. and Malkiel, B. (2005). Comparing modalities: Cognates as a case in point. *Across Languages and Cultures*, 6:173–193.

Shlesinger, M. and Ordan, N. (2012). More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target*, 24:1:43–60.

Slaats, S. and Martin, A. (2025). What's surprising about surprisal. *Computational Brain & Behavior*, 8:233–248.

Smith, N. and Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128:302–319.

Steiner, E. (2013). *A characterization of the resource based on shallow statistics*, pages 71–90. De Gruyter Mouton, Berlin, Boston.

Su, W. and Li, D. (2019). Identifying translation problems in English-Chinese sight translation: An eye-tracking experiment. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, 14(1):110–134.

Tang, F. (2018). *Explicitation in Consecutive Interpreting.* John Benjamins.

Teich, E. (2003). *Cross-Linguistic Variation in System und Text. A Methodology for the Investigation of Translations and Comparable Texts.* Mouton de Gruyter, Berlin.

Teich, E., Martînez Martînez, J., and Karakanta, A. (2020). Translation, information theory and cognition. In Alves, F. and Jakobsen, A. L., editors, *The Routledge Handbook of Translation and Cognition*, chapter 20, pages 360–375. Routledge, London.

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, 105(2):300–333.

Temperley, D. (2008). Dependency-length minimization in natural and artificial languages. *Journal of Quantitative Linguistics*, 15(3):256–282.

Tissi, B. (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive anaysis. *Linguistics*, pages 103–127.

Tommola, J. and Niemi, P. (1986). Mental load in simultaneous interpreting: An on-line pilot study. *Nordic research in text linguistics and discourse analysis*, pages 171–184.

Tourtouri, E. N., Delogu, F., and Crocker, M. W. (2021). Rational redundancy in referring expressions: Evidence from event-related potentials. *Cognitive Science*, 45.

Toury, G. (1980). In search of a theory of translation. *Porter Institute for Poetics and Semiotics*.

Toury, G. (1991). What are descriptive studies into translation likely to yield apart from isolated descriptions. *Translation Studies: The State of the Art*.

Toury, G. (1995). *Descriptive translation studies and beyond*. Benjamins translation library: 4. Benjamins.

Toury, G. (2012). *Descriptive Translation Studies – and beyond: Revised edition*. John Benjamins.

Tremblay, A. and Baayen, H. (2010). Holistic Processing of Regular Four-word Sequences: A Behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on Formulaic Language: Acquisition and Communication*, page 151–173.

Tremblay, A., Derwing, B., Libben, G., and Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2):569–613.

van Hell, J. G. and de Groot, A. M. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, 128(3):431–451. Bilingualism: Functional and neural perspectives.

Volansky, V., Ordan, N., and Wintner, S. (2015). On the features of translationese. *Digital Scholarship in the Humanities*, 30(1):98–118.

Wang, L. and Liu, H. (2013). Syntactic variations in Chinese–English code-switching. *Lingua*, 123:58–73.

Wang, Y. and Liu, H. (2017). The effects of genre on dependency distance and dependency direction. *Language Sciences*, 59:135–147.

Warren, T. and Gibson, E. (2002). The effects of discourse status on intuitive complexity: Implications for quantifying distance in a locality-based theory of linguistic complexity. *Cognition*, 85:79–112.

Wasow, T. (1997). End-weight from the speaker's perspective. *J Psycholinguist Res*, 26:347–361.

Weber, A. and Crocker, M. (2012). On the nature of semantic constraints on lexical access. *Journal of psycholinguistic research*, 41(3):195–214.

Wei, Y. (2022). Entropy as a measurement of cognitive load in translation. In Carl, M., Yamada, M., and Zou, L., editors, *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Workshop 1: Empirical Translation Process Research)*, pages 75–86. Association for Machine Translation in the Americas.

Wickens, C. D. (1976). The effects of divided attention on information processing in manual tracking. *J. Exp. Psychol. Hum. Percept. Perform.*, 2(1):1–13.

Wörrlein, M. (2007). *Der Simultandolmetschprozess – eine empirische Untersuchung.* Meidenbauer.

Xiao, R. and Dai, G. (2014). Lexical and grammatical properties of translational chinese: Translation universal hypotheses reevaluated from the chinese perspective. *Corpus Linguistics and Linguistic Theory*, 10.

Xiao, R. and Yue, M. (2010). Using corpora in translation studies: The state of the art. In *Contemporary Approaches to Corpus Linguistics*, chapter 14, pages 237–262. Continuum.

Xu, C. and Li, D. (2022). Exploring genre variation and simplification in interpreted language from comparable and intermodal perspectives. *Babel*, 68(5):742–770.

Xu, H. and Liu, K. (2023). Syntactic simplification in interpreted English: Dependency distance and direction measures. *Lingua*, 294:103607.

Yu, X., Falenska, A., and Kuhn, J. (2019). Dependency length minimization vs. word order constraints: An empirical study on 55 treebanks. In Chen, X. and Ferrer-i Cancho, R., editors, *Proceedings of the First Workshop on Quantitative Syntax*

(Quasy, SyntaxFest 2019), pages 89–97, Paris, France. Association for Computational Linguistics.

Yung, F., Scholman, M., Lapshinova-Koltunski, E., Pollkläsener, C., and Demberg, V. (2023). Investigating explicitation of discourse connectives in translation using automatic annotations. In Stoyanchev, S., Joty, S., Schlangen, D., Dusek, O., Kennington, C., and Alikhani, M., editors, *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 21–30, Prague, Czechia. Association for Computational Linguistics.

Zanetti, R. (1999). Relevance of anticipation and possible strategies in the simultaneous interpretation from English into Italian. *The Interpreters' Newsletter*, 9:79–98.

Zhang, R., Wang, X., Zhang, C., He, Z., Wu, H., Li, Z., Wang, H., Chen, Y., and Li, Q. (2021). BSTC: A large-scale Chinese-English speech translation dataset. In Wu, H., Cherry, C., Huang, L., He, Z., Liu, Q., Elbayad, M., Liberman, M., Wang, H., Ma, M., and Zhang, R., editors, *Proceedings of the Second Workshop on Automatic Simultaneous Translation*, pages 28–35, Online. Association for Computational Linguistics.

Łyda, A. and Gumul, E. (2002). Cohesion in interpreting. *PASE Papers in Language Studies: Proceedings of the Ninth Annual Conference of the Polish Association for the Study of English*, pages 349–356.

Appendices

Appendix A

Additional tables in the appendix

KLD values as basis for word cloud visualisation (Section 4.2)

Table A.1: KLD values as basis for word cloud visualisation: SI DE EN vs. ORG EN.

word	relative entropy	frequency	p-value
euh	0.03960763	0.04218045	0.00000000
hum	0.00698372	0.00375940	0.00001041
we	0.00596984	0.02483083	0.02644400
be	0.00460203	0.01135338	0.00547490
kosovo	0.00363414	0.00069549	0.00526050
't	0.00317923	0.00535714	0.01024100
need	0.00315935	0.00411654	0.00293900
're	0.00312006	0.00413534	0.00359710
can	0.00306121	0.00518797	0.00601240
going	0.00264991	0.00225564	0.00377160
policy	0.00259379	0.00146617	0.00915090
gentlemen	0.00239302	0.00065789	0.00002205
talking	0.00238054	0.00131579	0.00445210
external	0.00235728	0.00045113	0.00149750
ladies	0.00228686	0.00069549	0.00006145
gambling	0.00216084	0.00041353	0.04878800
guantánamo	0.00206262	0.00039474	0.01344400
citizens	0.00201887	0.00122180	0.04574100
something	0.00194038	0.00140977	0.00111650
europe	0.00192766	0.00195489	0.02778800
really	0.00154350	0.00184211	0.02499000
eulex	0.00137508	0.00026316	0.02612200

Continued on next page

word	relative entropy	frequency	p-value
freedom	0.00134374	0.00046992	0.01827800
austria	0.00127686	0.00024436	0.00402600
against	0.00126790	0.00114662	0.02897300
obviously	0.00120633	0.00054511	0.00863260
germany	0.00119494	0.00043233	0.01396700
transfer	0.00108042	0.00020677	0.03513100
president-in-office	0.00098340	0.00028195	0.00294290
shouldn	0.00098056	0.00050752	0.02400800
future	0.00092064	0.00095865	0.02827300
financial	0.00091838	0.00097744	0.04161500
border	0.00089963	0.00026316	0.00719300
certain	0.00088651	0.00069549	0.04223500
detainees	0.00088398	0.00016917	0.02963400
corruption	0.00088398	0.00016917	0.03999200
service	0.00087832	0.00060150	0.02732100
illegal	0.00084598	0.00030075	0.03595500
things	0.00079015	0.00107143	0.03889200
minority	0.00078576	0.00015038	0.00386270
merkel	0.00078576	0.00015038	0.03238400
austrian	0.00078576	0.00015038	0.01600500
resolution	0.00077615	0.00056391	0.03879100
justice	0.00077215	0.00028195	0.02172700
motion	0.00073520	0.00022556	0.03229600
'd	0.00073079	0.00129699	0.02472500
comes	0.00072943	0.00069549	0.02460500
serbia	0.00068754	0.00013158	0.04735900
guarantee	0.00062823	0.00024436	0.02654400
fundamental	0.00060057	0.00035714	0.02081300
step	0.00059692	0.00045113	0.01373200
arab	0.00058932	0.00011278	0.03103500
path	0.00058932	0.00011278	0.03185300
react	0.00058932	0.00011278	0.03180000
regards	0.00058932	0.00011278	0.01314600
applies	0.00058932	0.00011278	0.00969780
waiting	0.00058932	0.00011278	0.01457900
question	0.00057997	0.00112782	0.04002900
iran	0.00057580	0.00018797	0.02953600
epp	0.00057580	0.00018797	0.03160100
ready	0.00057580	0.00018797	0.01595500
civil	0.00050977	0.00030075	0.01437100

Continued on next page

word	relative entropy	frequency	p-value
quartet	0.00049110	0.00009398	0.04340400
fight	0.00049020	0.00020677	0.02791300

Table A.2: KLD values as basis for word cloud visualisation: ORG EN vs. SI DE EN.

word	relative entropy	frequency	p-value
i	0.00691985	0.01811679	0.00690470
on	0.00470029	0.00987273	0.00390810
our	0.00330152	0.00505390	0.00909200
my	0.00325766	0.00256892	0.00017945
british	0.00315225	0.00075557	0.00924450
feed	0.00256161	0.00050371	0.00322400
by	0.00238901	0.00359314	0.00645790
report	0.00213053	0.00213238	0.03347100
work	0.00196264	0.00201484	0.01405200
believe	0.00194179	0.00085631	0.00235510
was	0.00192274	0.00347560	0.00603850
irish	0.00187670	0.00067161	0.02146800
food	0.00187670	0.00067161	0.00891010
health	0.00186025	0.00075557	0.04231900
their	0.00181168	0.00315659	0.01448400
ireland	0.00172099	0.00075557	0.00540750
his	0.00155837	0.00104100	0.01494700
recognise	0.00153697	0.00030223	0.00016274
actually	0.00153360	0.00125928	0.03373800
had	0.00150735	0.00174620	0.02900300
tonight	0.00148367	0.00040297	0.00567320
at	0.00147261	0.00493636	0.03092100
enforcement	0.00145158	0.00028544	0.01617200
even	0.00144587	0.00097384	0.00036536
last	0.00141570	0.00125928	0.00112020
legislation	0.00139506	0.00100742	0.02270500
indeed	0.00139506	0.00100742	0.00525960
may	0.00137548	0.00083952	0.01725400
britain	0.00136619	0.00026865	0.00193260
relation	0.00133111	0.00036939	0.00186350
chamber	0.00128081	0.00025186	0.00672820
kingdom	0.00128081	0.00025186	0.00080461
accountability	0.00128081	0.00025186	0.02431600
record	0.00125550	0.00035260	0.01633600

Continued on next page

word	relative entropy	frequency	p-value
those	0.00125006	0.00230028	0.03952600
me	0.00120356	0.00112495	0.03329900
commitments	0.00119542	0.00023506	0.00406020
members	0.00117203	0.00065482	0.03983200
regulation	0.00112102	0.00063803	0.03306800
chair	0.00111003	0.00021827	0.00199080
face	0.00111003	0.00021827	0.00058163
requirements	0.00110579	0.00031902	0.00549030
issue	0.00104706	0.00112495	0.02917200
house	0.00104583	0.00105779	0.00529700
earlier	0.00103179	0.00030223	0.00857950
an	0.00103109	0.00357635	0.04809700
chain	0.00102464	0.00020148	0.02803100
fourty	0.00093926	0.00018469	0.01143400
industries	0.00093926	0.00018469	0.01155500
him	0.00093835	0.00033581	0.01081300
services	0.00093116	0.00065482	0.01258400
farmers	0.00090716	0.00053729	0.03105200
place	0.00086638	0.00082273	0.01612800
vote	0.00085776	0.00052050	0.04428800
commend	0.00085387	0.00016790	0.00639830
corporate	0.00085387	0.00016790	0.04773800
debate	0.00082656	0.00119212	0.04922500
systems	0.00077961	0.00033581	0.03656100
over	0.00077786	0.00120891	0.04992000
years	0.00077076	0.00122570	0.01619900
tribute	0.00076848	0.00015111	0.01105700
refused	0.00076848	0.00015111	0.00790590
yet	0.00075310	0.00058766	0.03011700
second	0.00073047	0.00041976	0.04168000
did	0.00071000	0.00057087	0.04028500

Table A.3: KLD values as basis for word cloud visualisation: SI EN DE vs. ORG DE.

word	relative entropy	frequency	p-value
euh	0.13025398	0.06117160	0.00000000
also	0.00991705	0.00419329	0.00000001
hum	0.00805377	0.00383335	0.00002856
hm	0.00674845	0.00185368	0.00000013
das	0.00655201	0.02075047	0.00005556

Continued on next page

word	relative entropy	frequency	p-value
da	0.00614845	0.00579501	0.00000408
wirklich	0.00542696	0.00338342	0.00013821
ja	0.00418684	0.00478719	0.00029221
aber	0.00404127	0.00689283	0.00138020
hier	0.00404115	0.00631693	0.00704150
jetzt	0.00400544	0.00503914	0.00013157
haben	0.00378935	0.01018627	0.00760990
dann	0.00346510	0.00453523	0.00154950
um	0.00297443	0.00467920	0.04203500
man	0.00290232	0.00484118	0.00726760
möchte	0.00288743	0.00311347	0.00583070
müssen	0.00252322	0.00464321	0.00819820
wichtig	0.00250197	0.00170971	0.00207710
einfach	0.00230153	0.00122379	0.00047647
ganz	0.00221240	0.00233960	0.00248550
denn	0.00212316	0.00232161	0.00066488
britische	0.00194232	0.00037794	0.02101800
so	0.00179292	0.00372537	0.02946700
sicherstellen	0.00175733	0.00034194	0.00005351
arbeiten	0.00168127	0.00077387	0.00182150
möchten	0.00164440	0.00050391	0.00301260
landwirte	0.00146507	0.00039593	0.00376450
doch	0.00144217	0.00194367	0.02030200
sektor	0.00138737	0.00026995	0.01804200
war	0.00135383	0.00196167	0.01476700
irland	0.00130700	0.00059390	0.03023400
weiter	0.00126804	0.00111581	0.00838090
gab	0.00119398	0.00059390	0.03523400
sagen	0.00117587	0.00228561	0.00483700
ganze	0.00111396	0.00079187	0.00962220
iren	0.00110989	0.00021596	0.03018500
uns	0.00105301	0.00349141	0.04692300
freue	0.00097261	0.00034194	0.02118500
’s	0.00090222	0.00061190	0.02715500
w	0.00069449	0.00026995	0.02616200
letzten	0.00067278	0.00091784	0.03198000
jobs	0.00064744	0.00012598	0.04439200

Table A.4: KLD values as basis for word cloud visualisation: ORG DE vs. SI EN DE.

word	relative entropy	frequency	p-value
der	0.01547948	0.03094852	0.00000001
in	0.00645700	0.01761912	0.00378820
kolleginnen	0.00396182	0.00108624	0.00000004
des	0.00388808	0.00592827	0.00033493
mit	0.00384489	0.00771412	0.00162740
von	0.00366855	0.00872671	0.00252830
liebe	0.00304861	0.00088372	0.00007686
kosovo	0.00304437	0.00058915	0.00926790
herren	0.00302183	0.00093895	0.00001344
als	0.00288532	0.00430812	0.00616070
europäischen	0.00284550	0.00307460	0.00100320
damen	0.00270989	0.00086531	0.00003464
wie	0.00248805	0.00502614	0.00731470
kollegen	0.00244188	0.00200678	0.00007986
den	0.00240571	0.01056779	0.02214200
guantánamo	0.00237841	0.00046027	0.00640800
sicht	0.00209300	0.00040504	0.00063025
sondern	0.00207660	0.00185949	0.00407480
meine	0.00202994	0.00156492	0.00076597
nach	0.00186599	0.00215406	0.00568870
is	0.00184959	0.00069961	0.00058182
einer	0.00184265	0.00244863	0.00649120
aus	0.00176983	0.00298255	0.01313700
bürgerinnen	0.00169029	0.00044186	0.00061831
herr	0.00165334	0.00373739	0.03356200
resolution	0.00151854	0.00040504	0.00451210
österreich	0.00142705	0.00027616	0.00369290
setzen	0.00142705	0.00027616	0.00034342
durch	0.00138515	0.00176744	0.04594400
politik	0.00131432	0.00071802	0.03503900
zur	0.00129158	0.00182267	0.02514700
unserer	0.00128455	0.00092054	0.00527330
dank	0.00119516	0.00244863	0.03508900
reden	0.00117719	0.00064438	0.01262000
bürger	0.00113421	0.00095736	0.04351400
keine	0.00112995	0.00156492	0.04605400
dienst	0.00108843	0.00036822	0.04762100
gegen	0.00106551	0.00110465	0.04869000

Continued on next page

word	relative entropy	frequency	p-value
am	0.00084152	0.00136240	0.03074400

Table A.5: Average surprisal for selected lexical verbs by delivery mode.

Verb and subcorpus	N	mean	SD	median	IQR	range
talk						
ORG EN impromptu	11	7.661818	3.202380	8.14	4.3150	9.91
SI DE EN (impromptu)	124	6.367661	3.552358	6.74	6.4025	14.14
SI ES EN (impromptu)	84	6.166190	3.495364	5.79	6.035	13.87
react						
SI DE EN (impromptu)	10	12.857	2.494171	12.33	4.18	6.04
SI ES EN (impromptu)	2	10.450	0.000000	10.45	0.00	0.00
believe						
ORG EN impromptu	17	9.585294	3.567039	11.07	5.170	10.87
SI DE EN (impromptu)	8	6.865000	1.412657	7.04	2.505	3.40
SI ES EN (impromptu)	28	7.743929	2.543324	7.55	2.855	10.87
sicherstellen						
ORG DE impromptu	1	12.84	NA	2.84	0.0000	0.00
SI EN DE (impromptu)	34	8.272941	3.730525	6.33	7.3425	11.83
arbeiten						
ORG DE impromptu	5	12.6520	2.852599	12.88	4.0800	6.70
SI EN DE (impromptu)	50	10.3242	2.303292	11.21	2.6925	10.97
sagen						
ORG DE impromptu	91	7.926264	2.841872	8.570	3.415	14.01
SI EN DE (impromptu)	252	7.200397	3.044108	7.705	4.620	14.00
geben						
ORG DE impromptu	64	7.817813	3.203734	9.13	6.1600	13.43
SI EN DE (impromptu)	224	7.636786	2.889122	9.12	4.3025	11.00
reden						
ORG DE impromptu	26	10.73885	3.790447	11.575	2.190	14.04
SI EN DE (impromptu)	12	10.60167	3.746826	11.560	1.845	11.00
setzen						
ORG DE impromptu	7	13.50571	1.0726269	13.07	0.445	3.08
SI EN DE (impromptu)	6	13.66333	0.5960425	13.42	0.000	1.46

Table A.7: 10 most frequent lexical verbs with surprisal 2.++.

	ORG EN	rf	SI DE EN	rf	ORG DE	rf	SI EN DE	rf
1	have	101	have	122	haben	34	haben	36
2	think	83	think	97	glaube	18	geht	28
3	is	48	need	66	geht	15	glaube	12
4	need	39	are	57	gibt	5	sagen	6
5	are	33	is	53	machen	5	gibt	5
6	's	27	talking	28	nehmen	5	fragen	4
7	want	16	's	27	gehört	4	bedanken	3
8	got	13	got	25	hat	4	bringen	3
9	thank	13	thank	25	sagen	4	danken	3
10	like	9	going	15	brauchen	3	gesagt	3

Table A.8: 10 most frequent nouns with surprisal 2.++.

	ORG EN	rf	SI DE EN	rf	ORG DE	rf	SI EN DE	rf
1	people	17	people	13	Frau	25	Frau	24
2	years	11	crisis	9	Damen	18	Herr	16
3	place	9	gambling	9	Herr	16	Prozent	16
4	time	8	lot	9	Ende	12	Jahren	12
5	trade	8	years	8	Fraktion	12	Kommission	10
6	crisis	7	Commission	7	Kommission	11	Bereich	9
7	Member	6	policy	7	Bereich	10	Beispiel	8
8	plan	6	entities	6	Kollegen	7	Ende	8
9	State	6	group	6	Prozent	7	Bericht	6
10	way	6	thing	6	Beispiel	6	Land	6

Table A.9: Universal dependency relations.

acl	clausal modifier of noun (adnominal clause)
acl:relcl	relative clause modifier
advcl	adverbial clause modifier
advmod	adverbial modifier
amod	adjectival modifier
appos	appositional modifier
aux	auxiliary
aux:pass	passive auxiliary
case	case marking
cc	coordinating conjunction
cc:preconj	preconjunct
ccomp	clausal complement
compound	compound
compound:prt	phrasal verb particle
conj	conjunct
cop	copula
csubj	clausal subject
csubj:pass	clausal passive subject
dep	unspecified dependency
det	determiner
det:predet	predeterminer
discourse	discourse element
expl	expletive
fixed	fixed multiword expression
flat	flat multiword expression
iobj	indirect object
mark	marker
nmod	nominal modifier
nmod:npmod	noun phrase as adverbial modifier
nmod:poss	possessive nominal modifier
nmod:tmod	temporal modifier
nsubj	nominal subject
nsubj:pass	passive nominal subject
nummod	numeric modifier
obj	object
obl	oblique nominal
obl:npmod	noun phrase as adverbial modifier
obl:tmod	temporal modifier
parataxis	parataxis
punct	punctuation
root	root
vocative	vocative
xcomp	open clausal complement

Table A.10: Average dependency length and frequency of each dependency relation for ORG EN vs. SI DE EN. Basis for normalisation: number of sentences. Only for raw frequency above 5.

Relation	av DL ENDEEN	ORG EN norm	SI DE EN norm	normDiff
acl	2.6457	19.3263	11.2538	8.0726
acl:relcl	3.8860	33.4898	22.2344	11.2554
advcl	7.4242	31.7504	24.2830	7.4674
advmod	2.3413	108.7521	96.2305	12.5215
amod	1.2704	102.3192	87.3805	14.9387
appos	2.7376	3.7272	2.3491	1.3781
aux	1.6966	68.4428	67.9596	0.4833
aux:pass	1.2535	16.8415	18.0279	-1.1863
case	2.0508	188.4318	150.9970	37.4348
cc	3.5984	77.9404	63.6711	14.2692
cc:preconj	1.9474	0.6902	0.3551	0.3351
ccomp	5.8452	29.2104	26.1677	3.0427
compound	1.1918	56.9851	52.0076	4.9774
compound:prt	1.1377	6.5986	6.9653	-0.3667
conj	5.4126	47.6256	35.2363	12.3893
cop	2.1566	43.1806	41.9011	1.2794
csubj	5.0087	3.2027	3.0866	0.1161
det	1.5619	176.7532	145.2882	31.4650
det:predet	2.3855	2.6505	1.9120	0.7384
discourse	3.3081	28.6858	53.9743	-25.2885
expl	1.7411	9.3871	9.9153	-0.5282
fixed	1.0815	7.4544	6.0366	1.4178
flat	1.2462	10.4914	7.8121	2.6794
iobj	1.2308	1.5737	1.6389	-0.0652
mark	2.8288	98.5643	84.1027	14.4616
nmod	3.1798	88.2109	70.2540	17.9569
nmod:npmod	1.5714	0.7178	0.4370	0.2808
nmod:poss	1.3863	23.3573	12.8107	10.5466
nmod:tmod	2.9545	0.5246	0.0819	0.4426
nsubj	2.6499	162.2584	148.2928	13.9656
nsubj:pass	3.9631	13.7493	14.5042	-0.7549
nummod	1.6481	13.4456	9.7514	3.6942
obj	2.2615	103.8377	78.8309	25.0067
obl	4.4188	92.5179	75.5531	16.9648
obl:npmod	2.0902	1.5461	1.8028	-0.2567
obl:tmod	4.2862	4.8040	2.5949	2.2091
parataxis	12.0825	2.1811	3.1412	-0.9601
vocative	3.9123	0.8835	0.6829	0.2006
xcomp	2.3660	34.3733	35.0178	-0.6445

Table A.11: Average dependency length and frequency of each dependency relation for ORG EN vs. SI DE EN. Basis for normalisation: number of relations. Only for raw frequency above 5.

Relation	av DL ENDEEN	ORG EN norm	SI DE EN norm	normDiff
acl	2.6457	1.0083	0.6718	0.3365
acl:relcl	3.8860	1.7472	1.3273	0.4199
advcl	7.4242	1.6565	1.4496	0.2069
advmod	2.3413	5.6738	5.7447	-0.0709
amod	1.2704	5.3382	5.2164	0.1218
appos	2.7376	0.1945	0.1402	0.0542
aux	1.6966	3.5708	4.0570	-0.4862
aux:pass	1.2535	0.8787	1.0762	-0.1976
case	2.0508	9.8309	9.0141	0.8168
cc	3.5984	4.0663	3.8010	0.2653
cc:preconj	1.9474	0.0360	0.0212	0.0148
ccomp	5.8452	1.5240	1.5621	-0.0382
compound	1.1918	2.9730	3.1047	-0.1317
compound:prt	1.1377	0.3443	0.4158	-0.0715
conj	5.4126	2.4847	2.1035	0.3812
cop	2.1566	2.2528	2.5014	-0.2486
csubj	5.0087	0.1671	0.1843	-0.0172
det	1.5619	9.2216	8.6733	0.5483
det:predet	2.3855	0.1383	0.1141	0.0241
discourse	3.3081	1.4966	3.2221	-1.7255
expl	1.7411	0.4897	0.5919	-0.1022
fixed	1.0815	0.3889	0.3604	0.0285
flat	1.2462	0.5474	0.4664	0.0810
iobj	1.2308	0.0821	0.0978	-0.0157
mark	2.8288	5.1423	5.0207	0.1216
nmod	3.1798	4.6022	4.1940	0.4082
nmod:npmod	1.5714	0.0375	0.0261	0.0114
nmod:poss	1.3863	1.2186	0.7648	0.4538
nmod:tmod	2.9545	0.0274	0.0049	0.0225
nsubj	2.6499	8.4654	8.8527	-0.3873
nsubj:pass	3.9631	0.7173	0.8659	-0.1485
nummod	1.6481	0.7015	0.5821	0.1194
obj	2.2615	5.4174	4.7060	0.7114
obl	4.4188	4.8269	4.5103	0.3165
obl:npmod	2.0902	0.0807	0.1076	-0.0270
obl:tmod	4.2862	0.2506	0.1549	0.0957
parataxis	12.0825	0.1138	0.1875	-0.0737
vocative	3.9123	0.0461	0.0408	0.0053
xcomp	2.3660	1.7933	2.0905	-0.2971

Table A.12: Average dependency length and frequency of each dependency relation for ORG EN vs. SI ES EN. Basis for normalisation: number of sentences. Only for raw frequency above 5.

Relation	av DL ENESEN	ORG EN norm	SI ES EN norm	normDiff
acl	2.5823	19.3263	15.3771	3.9492
acl:relcl	3.9099	33.4898	29.4538	4.0359
advcl	7.3202	31.7504	27.6983	4.0521
advmod	2.2888	108.7521	89.0117	19.7404
amod	1.2624	102.3192	106.1118	-3.7927
appos	3.0364	3.7272	4.5514	-0.8241
aux	1.6442	68.4428	66.2224	2.2205
aux:pass	1.2225	16.8415	13.3290	3.5125
case	2.0497	188.4318	176.2354	12.1964
cc	3.4073	77.9404	70.4811	7.4592
cc:preconj	2.1707	0.6902	0.5202	0.1701
ccomp	5.7246	29.2104	25.2276	3.9828
compound	1.1764	56.9851	49.1873	7.7978
compound:prt	1.0919	6.5986	6.7295	-0.1310
conj	5.1682	47.6256	42.1001	5.5255
cop	2.1429	43.1806	40.1495	3.0310
csubj	5.3084	3.2027	3.6086	-0.4059
det	1.5622	176.7532	170.6112	6.1420
det:predet	2.3846	2.6505	1.9506	0.6999
discourse	3.1888	28.6858	58.1599	-29.4741
expl	1.7405	9.3871	10.3706	-0.9835
fixed	1.0782	7.4544	7.4447	0.0097
flat	1.2388	10.4914	8.6151	1.8764
iobj	1.1386	1.5737	1.4304	0.1433
mark	2.6780	98.5643	90.9298	7.6346
nmod	3.1767	88.2109	90.6372	-2.4263
nmod:npmod	1.5405	0.7178	0.3576	0.3602
nmod:poss	1.3684	23.3573	15.7347	7.6225
nmod:tmod	3.0385	0.5246	0.2276	0.2970
nsubj	2.6162	162.2584	152.8283	9.4301
nsubj:pass	3.9280	13.7493	9.5579	4.1914
nummod	1.6054	13.4456	9.4603	3.9853
obj	2.2888	103.8377	99.7399	4.0977
obl	4.4214	92.5179	81.3719	11.1460
obl:npmod	2.3200	1.5461	1.4304	0.1157
obl:tmod	4.1970	4.8040	2.9259	1.8781
parataxis	11.9231	2.1811	2.5033	-0.3221
vocative	3.7708	0.8835	0.5202	0.3633
xcomp	2.3549	34.3733	40.7672	-6.3940

Table A.13: Average dependency length and frequency of each dependency relation for ORG EN vs. SI ES EN. Basis for normalisation: number of relations. Only for raw frequency above 5.

Relation	av DL ENESEN	ORG EN norm	SI ES EN norm	normDiff
acl	2.5823	1.0083	0.8430	0.1653
acl:relcl	3.9099	1.7472	1.6147	0.1325
advcl	7.3202	1.6565	1.5185	0.1380
advmod	2.2888	5.6738	4.8798	0.7940
amod	1.2624	5.3382	5.8172	-0.4790
appos	3.0364	0.1945	0.2495	-0.0551
aux	1.6442	3.5708	3.6304	-0.0596
aux:pass	1.2225	0.8787	0.7307	0.1479
case	2.0497	9.8309	9.6616	0.1693
cc	3.4073	4.0663	3.8639	0.2024
cc:preconj	2.1707	0.0360	0.0285	0.0075
ccomp	5.7246	1.5240	1.3830	0.1409
compound	1.1764	2.9730	2.6965	0.2765
compound:prt	1.0919	0.3443	0.3689	-0.0247
conj	5.1682	2.4847	2.3080	0.1767
cop	2.1429	2.2528	2.2011	0.0518
csubj	5.3084	0.1671	0.1978	-0.0307
det	1.5622	9.2216	9.3532	-0.1316
det:predet	2.3846	0.1383	0.1069	0.0313
discourse	3.1888	1.4966	3.1884	-1.6918
expl	1.7405	0.4897	0.5685	-0.0788
fixed	1.0782	0.3889	0.4081	-0.0192
flat	1.2388	0.5474	0.4723	0.0751
iobj	1.1386	0.0821	0.0784	0.0037
mark	2.6780	5.1423	4.9849	0.1574
nmod	3.1767	4.6022	4.9689	-0.3667
nmod:npmod	1.5405	0.0375	0.0196	0.0178
nmod:poss	1.3684	1.2186	0.8626	0.3560
nmod:tmod	3.0385	0.0274	0.0125	0.0149
nsubj	2.6162	8.4654	8.3783	0.0870
nsubj:pass	3.9280	0.7173	0.5240	0.1934
nummod	1.6054	0.7015	0.5186	0.1829
obj	2.2888	5.4174	5.4679	-0.0505
obl	4.4214	4.8269	4.4610	0.3659
obl:npmod	2.3200	0.0807	0.0784	0.0022
obl:tmod	4.1970	0.2506	0.1604	0.0902
parataxis	11.9231	0.1138	0.1372	-0.0234
vocative	3.7708	0.0461	0.0285	0.0176
xcomp	2.3549	1.7933	2.2349	-0.4416

Table A.14: Average dependency length and frequency of each dependency relation for the German subset. Basis for normalisation: number of sentences. Only for raw frequency above 5.

Relation	av DL DE	ORG DE norm	SI EN DE norm	normDiff
acl	6.8753	15.6351	13.8862	1.7489
advcl	8.5900	17.2778	14.9902	2.2876
advmod	2.9694	203.6961	233.6114	-29.9153
amod	1.1989	102.2881	82.5810	19.7071
appos	1.3773	9.0642	5.8145	3.2497
aux	3.7561	60.2816	61.4573	-1.1757
aux:pass	2.0841	17.9231	15.6526	2.2705
case	1.8876	135.4356	92.4190	43.0166
cc	4.0358	68.2605	57.4828	10.7777
ccomp	8.3664	23.6140	21.7615	1.8524
compound	1.0547	8.4189	6.4033	2.0156
compound:prt	5.1720	3.9014	2.8705	1.0310
conj	5.7782	58.2869	35.8685	22.4184
cop	2.6843	28.1608	25.2453	2.9154
csubj	8.6075	3.4908	1.6438	1.8470
csubj:pass	8.7105	0.5867	0.4416	0.1451
dep	1.7184	1.4960	1.2758	0.2203
det	1.3715	216.3684	157.5319	58.8365
det:poss	1.2690	13.7577	10.7458	3.0119
expl	2.6276	4.4588	4.4406	0.0182
fixed	1.3571	0.4693	0.2944	0.1749
flat	1.3619	8.5656	4.5633	4.0023
iobj	3.3724	10.9710	8.6114	2.3596
mark	5.1024	56.5268	48.3072	8.2197
nmod	2.7435	102.3761	59.4946	42.8815
nmod:poss	1.1818	0.3520	0.2453	0.1067
nsubj	3.8695	152.1267	139.2787	12.8480
nsubj:pass	4.4600´	15.9871	14.2051	1.7820
nummod	1.2113	5.7788	5.6183	0.1606
obj	3.3920	97.1546	87.0461	10.1085
obl	4.1154	75.8287	53.9745	21.8542
parataxis	6.6944	1.1440	0.8096	0.3344
xcomp	5.8738	17.4245	14.0088	3.4156

Table A.15: Average dependency length and frequency of each dependency relation for the German subset. Basis for normalisation: number of dependency relations. Only for raw frequency above 5.

Relation	av DL DE	ORG DE norm	SI EN DE norm	normDiff
acl	6.8753	0.8992	0.9347	-0.0356
advcl	8.5900	0.9936	1.0090	-0.0154
advmod	2.9694	11.7145	15.7251	-4.0106
amod	1.1989	5.8826	5.5588	0.3238
appos	1.3773	0.5213	0.3914	0.1299
aux	3.7561	3.4668	4.1369	-0.6701
aux:pass	2.0841	1.0308	1.0536	-0.0229
case	1.8876	7.7889	6.2210	1.5679
cc	4.0358	3.9256	3.8693	0.0563
ccomp	8.3664	1.3580	1.4648	-0.1068
compound	1.0547	0.4842	0.4310	0.0531
compound:prt	5.1720	0.2244	0.1932	0.0312
conj	5.7782	3.3521	2.4144	0.9376
cop	2.6843	1.6195	1.6993	-0.0798
csubj	8.6075	0.2008	0.1106	0.0901
csubj:pass	8.7105	0.0337	0.0297	0.0040
dep	1.7184	0.0860	0.0859	0.0002
det	1.3715	12.4433	10.6039	1.8393
det:poss	1.2690	0.7912	0.7233	0.0679
expl	2.6276	0.2564	0.2989	-0.0425
fixed	1.3571	0.0270	0.0198	0.0072
flat	1.3619	0.4926	0.3072	0.1854
iobj	3.3724	0.6309	0.5797	0.0513
mark	5.1024	3.2508	3.2517	-0.0009
nmod	2.7435	5.8876	4.0048	1.8829
nmod:poss	1.1818	0.0202	0.0165	0.0037
nsubj	3.8695	8.7488	9.3753	-0.6265
nsubj:pass	4.4600	0.9194	0.9562	-0.0368
nummod	1.2113	0.3323	0.3782	-0.0458
obj	3.3920	5.5873	5.8593	-0.2720
obl	4.1154	4.3609	3.6332	0.7277
parataxis	6.6944	0.0658	0.0545	0.0113
xcomp	5.8738	1.0021	0.9430	0.0591

Table A.16: Average dependency length and differences in average dependency length for ORG EN vs. SI DE EN. Only for raw frequency above 5.

Relation	ORG EN avDL	SI DE EN avDL	avDLDiff
acl	2.7343	2.4951	0.2391
acl:relcl	3.9308	3.8194	0.1113
advcl	7.2670	7.6277	-0.3607
advmod	2.2668	2.4246	-0.1578
amod	1.2744	1.2657	0.0087
appos	2.7778	2.6744	0.1034
aux	1.6688	1.7243	-0.0555
aux:pass	1.2426	1.2636	-0.0210
case	2.0560	2.0443	0.0117
cc	3.5285	3.6830	-0.1545
cc:preconj	2.0000	1.8462	0.1538
ccomp	5.8043	5.8904	-0.0860
compound	1.1768	1.2080	-0.0311
compound:prt	1.1255	1.1490	-0.0235
conj	5.3438	5.5047	-0.1609
cop	2.1566	2.1565	0.0002
csubj	5.7155	4.2832	1.4323
det	1.5439	1.5836	-0.0397
det:predet	2.3333	2.4571	-0.1238
discourse	3.5582	3.1766	0.3816
expl	1.7529	1.7300	0.0229
fixed	1.1000	1.0588	0.0412
flat	1.2316	1.2657	-0.0342
iobj	1.1930	1.2667	-0.0737
mark	2.7975	2.8652	-0.0677
nmod	3.2135	3.1380	0.0754
nmod:npmod	1.4231	1.8125	-0.3894
nmod:poss	1.3842	1.3902	-0.0060
nmod:tmod	2.7368	4.3333	-1.5965
nsubj	2.6546	2.6449	0.0097
nsubj:pass	3.8715	4.0490	-0.1775
nummod	1.6735	1.6134	0.0601
obj	2.2582	2.2658	-0.0076
obl	4.4551	4.3749	0.0802
obl:npmod	2.0000	2.1667	-0.1667
obl:tmod	4.5920	3.7263	0.8656
parataxis	11.5063	12.4783	-0.9719
vocative	3.7500	4.1200	-0.3700
xcomp	2.3783	2.3541	0.0242

Table A.17: Average dependency length and differences in average dependency length for ORG EN vs. SI ES EN. Only for raw frequency above 5.

Relation	ORG EN avDL	SI ES EN avDL	avDLDiff
acl	2.7343	2.3573	0.3770
acl:relcl	3.9308	3.8819	0.0489
advcl	7.2670	7.3920	-0.1251
advmod	2.2668	2.3203	-0.0535
amod	1.2744	1.2488	0.0256
appos	2.7778	3.2857	-0.5079
aux	1.6688	1.6141	0.0547
aux:pass	1.2426	1.1927	0.0499
case	2.0560	2.0419	0.0141
cc	3.5285	3.2495	0.2790
cc:preconj	2.0000	2.4375	-0.4375
ccomp	5.8043	5.6160	0.1884
compound	1.1768	1.1758	0.0010
compound:prt	1.1255	1.0531	0.0724
conj	5.3438	4.9344	0.4094
cop	2.1566	2.1255	0.0311
csubj	5.7155	4.8829	0.8326
det	1.5439	1.5846	-0.0407
det:predet	2.3333	2.4667	-0.1333
discourse	3.5582	2.9743	0.5839
expl	1.7529	1.7273	0.0257
fixed	1.1000	1.0524	0.0476
flat	1.2316	1.2491	-0.0175
iobj	1.1930	1.0682	0.1248
mark	2.7975	2.5256	0.2719
nmod	3.2135	3.1345	0.0790
nmod:npmod	1.4231	1.8182	-0.3951
nmod:poss	1.3842	1.3409	0.0433
nmod:tmod	2.7368	3.8571	-1.1203
nsubj	2.6546	2.5682	0.0864
nsubj:pass	3.8715	4.0238	-0.1523
nummod	1.6735	1.4914	0.1821
obj	2.2582	2.3263	-0.0681
obl	4.4551	4.3763	0.0787
obl:npmod	2.0000	2.7273	-0.7273
obl:tmod	4.5920	3.4333	1.1586
parataxis	11.5063	12.3506	-0.8443
vocative	3.7500	3.8125	-0.0625
xcomp	2.3783	2.3317	0.0466

Table A.18: Average dependency length and differences in average dependency length for the German subset. Only for raw frequency above 5.

Relation	ORG DE avDL	SI EN DE avDL	avDLDiff
acl	7.3527	6.4258	0.9269
advcl	8.9881	8.2062	0.7819
advmod	3.0230	2.9303	0.0928
amod	1.1844	1.2139	-0.0295
appos	1.3754	1.3797	-0.0043
aux	3.8302	3.6954	0.1348
aux:pass	2.2128	1.9608	0.2520
case	1.8659	1.9143	-0.0483
cc	3.7585	4.3111	-0.5527
ccomp	9.0447	7.7508	1.2939
compound	1.0592	1.0498	0.0094
compound:prt	5.4511	4.8547	0.5964
conj	5.7076	5.8741	-0.1665
cop	2.7323	2.6395	0.0928
csubj	8.8151	8.2388	0.5763
csubj:pass	9.2500	8.1111	1.1389
dep	1.8039	1.6346	0.1693
det	1.3661	1.3778	-0.0118
det:poss	1.2537	1.2854	-0.0317
expl	2.6711	2.5912	0.0799
fixed	1.5000	1.1667	0.3333
flat	1.4555	1.2151	0.2404
iobj	3.5053	3.2308	0.2746
mark	5.4650	4.7476	0.7174
nmod	2.7181	2.7802	-0.0622
nmod:poss	1.3333	1.0000	0.3333
nsubj	4.0262	3.7263	0.3000
nsubj:pass	4.7872	4.1520	0.6352
nummod	1.1726	1.2445	-0.0720
obj	3.7129	3.0924	0.6204
obl	4.3629	3.8245	0.5383
parataxis	6.8205	6.5455	0.2751
xcomp	6.3620	5.3660	0.9959