

Article

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

Volker Gast ^{1,*}  and Robert Borges ² ¹ Department of English and American Studies, Friedrich Schiller University, 07743 Jena, Germany² Institute of Slavic Studies, Polish Academy of Sciences, 00-337 Warsaw, Poland

* Correspondence: volker.gast@uni-jena.de

Abstract: This study investigates the distributions of word classes in English speeches made in the European Parliament and their German (written) translations and simultaneous interpretations. For comparison, a sample of original German speeches and a selection of political interviews are used. The study is motivated by the intention to understand the relationship between the type of mediation and communicative modes: mediated spoken language is compared to unmediated spoken language and to mediated written language. The results show that the interpretations exhibit a less nominal style than the translations, in this respect resembling unplanned spoken conversation. Other quantitative findings, such as a high frequency of adverbs, also point to a register effect, but interpretations have a hybrid status and can be located somewhere in the middle, between the register of the source text (parliamentary speech) and unplanned spoken discourse. The results are discussed against the background of the mechanisms that presumably underlie the choices made by translators (processing, register and strategies).

Keywords: translation; interpretation; triangular corpus; bilingual cognition; parts of speech



Citation: Gast, Volker, and Robert Borges. 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: 39. <https://doi.org/10.3390/languages8010039>

Academic Editor: Jeanine Treffers-Daller

Received: 7 December 2022

Revised: 17 January 2023

Accepted: 24 January 2023

Published: 29 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

It is well known that translated language differs from non-translated (original) language in systematic ways. Such differences were approached in terms of “universals of translation” (Baker 1993, 1996). While the exact source of any differences (linguistic-cognitive and social) remains a matter of dispute (see, for instance, contributions to Mauranen and Kujamäki 2004), there is a broad consensus that there are general properties typically exhibited by translated language. Among the “classics” of translation universals are simplification, explicitation, normalization and leveling out (Baker 1993, 1996; Laviosa-Braithwaite 2001).

More recently, the search for generalizations was extended from written translation to simultaneous interpreting (Gumul 2006; Defrancq et al. 2015; Lapshinova-Koltunski et al. 2021, 2022). Simultaneous interpreting—according to De Groot (2011, p. 222), “the most complex of all forms of linguistic behavior”—takes place under time pressure and cognitive strain, implying parallel brain activity at various levels (De Groot 1997; Christofels and de Groot 2005; Liu 2001; Liu et al. 2004, 2020). In his “Effort model”, Gile (2009) distinguishes a “Listening effort”, a “Memory Effort”, a “Production Effort” and a “Coordination Effort”. The high cognitive load during interpreting is reflected, among other things, in non-fluencies (Dayter 2021). Given the cognitive effort associated with simultaneous interpreting, interpretations can be expected to differ from both the original texts and the written translations in systematic ways. For example, in a study of discourse connectives, Lapshinova-Koltunski et al. (2022, p. 15) found that “equivalence and implicitation are

more frequently used in interpreting than translation, as these strategies facilitate cognitive processing in high-time-pressure situations”.

The time factor is obviously eminently important in interpreting. Interpreters often compress the source texts, ideally minimizing the loss of information. Consider (1), which deals with the appointment of the President of the European Council and the High Representative:

- (1) We should start by defining clearly their scope and nature, and then we should establish the qualities and experience of the people needed to fill them.
- (2) Wir sollten damit beginnen, die Art und den Umfang der Rolle deutlich zu definieren, und dann sollten wir die Fähigkeiten und Erfahrung anführen, die für diese Positionen erforderlich sind.

In sentence (1), the *scope* and *nature* of the position are under discussion (in German, *Art* and *Umfang*) and the *qualities* and *experience* of the candidates for the job (in German, *Fähigkeiten* and *Erfahrung*, cf. (2)). The interpretation of this passage summarizes *scope* and *nature* under *Zielsetzungen* ‘objectives’ and *qualities* and *experience* under *Erwartungen* ‘expectations’:¹

- (3) Wir sollten die Zielsetzung formulieren und dann auch die entsprechenden Erwartungen an die Amtsträger.

Beyond general tendencies or universals of translation and interpreting, recurring patterns have been observed in the relationship between originals and translations from specific language pairs. For example, it was observed that there are systematic shifts in the frequency distributions of word classes in English–German translations (Čulo et al. 2008; Steiner 2001; Hansen and Hansen-Schirra 2012; Steiner 2012). Čulo et al. (2008) found, among other things, that in a parallel corpus of corporate communication, there were “more verb to x alignments for English–German, but fewer x to noun alignments and more noun to x alignments for German–English” (Čulo et al. 2008, p. 50). In other words, translators tended to render verbal structures in English originals in a non-verbal way in the German translations. Conversely, nouns in the German originals were often rendered differently in English, e.g., with adjectives and verbs. This finding was confirmed by Serbina et al. (2017) in an experimental study. Serbina et al. (2017) provide a process-based analysis of translation shifts, using keystroke logging and eye-tracking data. According to their results, verb–noun shifts are associated with longer total fixation duration values, and both verb–noun shifts and noun–verb shifts show a higher number of fixations. The authors interpret these findings in terms of the grammatical complexity of parts of speech.

Part-of-speech distributions were also studied in interpreting. Shlesinger (2008) compares the distributions of parts of speech in English–Hebrew written translations and interpretations of the same source texts, finding notable differences in the distributions of pronouns and adjectives as well as, to some extent, function words, such as prepositions, conjunctions and copulas. In general, the interpretations seem to exhibit features of spoken language, though “it may also be argued that interpreted discourse displays features which set it apart; i.e., features of *interpretese*” (Shlesinger 2008, p. 250). Shlesinger and Ordan (2012) provide additional corpus evidence confirming that interpreted language exhibits features of spoken language. Pronouns and adverbs are more frequent in simultaneous interpretations than in written translation. Moreover, Shlesinger and Ordan (2012) show that proper nouns are severely underrepresented in simultaneous interpretations, which is interpreted as reflecting the oral nature of interpreting.

Similar results were obtained by Przybyl et al. (2022) for English–German translation and interpreting. The authors identify several features typical of interpreting—hesitation markers, intensifiers and a verbal style, with characteristically general verbs—as well as features that seem to be language specific, e.g., discourse particles, deictic adverbs and conjunctions. Written translation exhibits a more nominal style in German, and pronouns are typical of simultaneous interpretation in both English and German. Much in the spirit

of Shlesinger and Ordan (2012), the authors conclude that “interpreting seems to be more spoken than originals” (Przybyl et al. 2022, p. 201).

Dayter (2018) deals with the distributions of POS-tags in English and Russian. She finds that in both English–Russian and Russian–English interpreting, the ratio of nouns to verbs changes in the direction of the source language. The English interpretations of Russian texts exhibit a higher noun–verb ratio than the English originals, and the Russian interpretations of English texts exhibit a lower noun–verb ratio than the Russian originals. Dayter (2018) interprets these asymmetries as reflecting a *shining-through* effect (Teich 2003).

There are three major types of explanations for differences in part-of-speech distributions as described above. The first type of explanation could be called the “processing theory”. According to this line of reasoning, translation and interpreting imply routes of (bilingual) cognition that differ from those of monolingual language production (e.g., Grosjean 2001; De Groot 2011; Halverson 2014). Such ideas were pursued by Steiner (2001), providing a point of reference for other studies, e.g., Serbina et al. (2017). Second, the differences between original texts and translated language could be related to the fact that the text types have different register characteristics. This is particularly plausible in the simultaneous interpretation of formal texts, which introduces more (conceptionally) oral features (cf. Zeller Mayer 1990 on English–Hebrew translation). Note that an “equalizing” register effect was also claimed more generally, in the sense that shifts can be observed in the other direction as well (oral to literate; cf. Pym 2007; Shlesinger 2008). Finally, systematic differences between originals and translated texts could be due to strategies employed by translators and interpreters (Dayter 2021).

The three perspectives on differences between original and translated language are of course not mutually exclusive and in fact hard to keep apart. While the study of register originally focused on properties of texts (cf. Biber’s 1988 dimensions of register variation), there is a growing awareness that there is a cognitive dimension to register variation, e.g., insofar as registers are assumed to be represented as probabilistic grammars in speakers’ minds (Szmrecsanyi and Engel 2022). This means that register variation could have a cognitive basis, e.g., insofar as it involves cognitive routines associated with specific communicative situations. Interpreting strategies, too, can be regarded as cognitive routines, though they involve control and potentially training (see Dayter 2018 for a discussion of the term “strategy”).

Teasing apart the determinants of register variation in mediated language is a non-trivial matter and certainly requires experimental approaches (cf. Serbina et al. 2017). However, corpus studies can also be informative, as the examples mentioned above show. Before any far-reaching theories can be formulated, we first need to gain a better idea of the space of variation. The present study is intended to make a contribution to this agenda by comparing the distributions of the parts of speech in interpretations of political speeches to the distributions in the relevant originals and written translations, as well as a reference corpus of political interviews. It is based on a triangular corpus containing English political speeches as well as their German translations and interpretations. The approach is largely exploratory, following the example of the studies mentioned above. The main objective of the study is to identify linguistic features that are characteristic of interpreting. On the basis of the results obtained in studies of this type, the bigger questions can be addressed: Is interpreting “just” mediated spoken language? Or is it “an extreme case of translation” (Shlesinger and Ordan 2012, p. 54)? Are the effects of mediation and the mode of production additive, or is interpreting *sui generis*? The present study does not provide conclusive answers to these questions, but it aims to gather evidence functioning as a basis for follow-up research.

Section 2 describes the data used for the present study. The results are presented in Section 3. Section 4 contains the Discussion and the Conclusions.

2. Materials and Methods

Recently, there have been a number of corpus building initiatives for interpreted language, e.g., the European Parliament Interpreting Corpus (EPIC, cf. [Bendazzoli and Sandrelli 2005](#); and EPIC-UdS; cf. [Przybyl et al. 2022](#)), the Polish Interpreting Corpus (PINC, cf. [Chmiel et al. 2022](#)) and the Spanish PETIMOD corpus ([Pastor and Rodas 2022](#); see also other contributions to [Kajzer-Wietrzny et al. 2022a](#)). Moreover, there are corpora that contain both interpreting and translation data, e.g., the European Parliament Interpreting and Translation Corpus (EPTIC, cf. [Bernardini et al. 2018](#); [Ferraresi and Bernardini 2019](#)). These corpora allow for a broad range of applications, e.g., the study of interpreters' behavior during interpreting (e.g., [Defrancq 2015](#); [Defrancq and Plevoets 2018](#)), and a comparison of translations and interpretations. The present study was carried out in that spirit, based on a "triangular" corpus of political speeches made in the European Parliament. The corpus contains the original speeches in the form of "verbatim reports", which are precise renderings of the speeches, as well as the simultaneous interpretations of the speeches, and the written translations of the verbatim reports (see, for instance, [Kajzer-Wietrzny et al. 2022b](#) for details). Most of the speeches are read out from a script, though some speakers deliver their speeches impromptu (often holding a manuscript as well).

The triangular corpus underlying the present study consists of 34,458 mp4 files with speeches made in the European Parliament between September 2009 and May 2010 (97 days, approx. 875 h of recording).² The relevant mp4 files were obtained from the website of the European Parliament. They contain the video signal as well as audio tracks for 22 languages, the language in which the speech was made as well as 21 other languages with simultaneous interpretations. The audio tracks were transcribed automatically, with acceptable, but not perfect, results. The data were time aligned and stored in Elan Annotation Format ([Wittenburg et al. 2006](#)).

In order to create a sample, we extracted all English speeches of at least 30 s of length (net speaking time) and filtered according to the quality of the transcription. The transcription software provides a confidence estimate per chunk of transcribed speech, and we only used speeches with an average confidence score of at least 0.5. This left us with a sample of 172 speeches. Fifty of these speeches were selected randomly and corrected manually using Hexatomic.³ Our manual corrections were restricted to obvious errors made by the speech recognition system. After processing the texts as described below, the corpus consisted of 33,850 tokens of English originals (29,794 tokens after removing unanalyzed tags such as punctuation), 33,531 tokens of German translations (29,088 tokens after filtering) and 27,136 words of German interpretations (26,582 tokens after filtering).

Figure A1 in Appendix A.2 shows density plots for the number of words per speech in our sample. The distributions are bimodal because the speeches differ systematically in length depending on the speaker. Longer speeches (>1500 words) are normally given by representatives of the executive, the European Commission, while the shorter speeches are delivered by members of the Parliament. The data do not show any indication of a length effect on the questions investigated in this study.

The speeches are listed in Table A1 in Appendix A.1. A total of 22 speeches were given by MEPs representing a country in which English is an official language, i.e., either the UK (19) or Ireland (3). The nationalities of the other speakers can be found in Table A1. We take it that the accent of a speaker does not have a significant influence on the process of interpreting, as the speakers are highly proficient (English is a working language in the European Parliament) and articulate very clearly. Moreover, the interpreters are used to working under these conditions (non-native speech). From a grammatical point of view, the speeches are impeccable.

As for the mode of delivery, the majority of speeches, native or non-native, are read out from a script ("manuscript speaking"). Seven of the sample speeches seem to be given extemporaneously, with at least three of these speakers holding, and occasionally looking at, a script.⁴ There is no perceptible difference in the speed of delivery, or the accu-

racy of articulation, between the read-out speeches and the speeches that were delivered extemporaneously.

As we wanted to take the identity of the interpreters into account, and because that information is not publicly available (if at all), we created sound embeddings (with 512 dimensions) of the speech signals with the Python package `resemblyzer`.⁵ The embeddings allow us to identify degrees of similarity between voices. By applying k-means clustering, evaluating the cluster solutions using the silhouette method, we determined 21 interpreters for the sample of 50 speeches. Manual checks of the interpreters assigned to the speeches showed the method to be robust, but we cannot exclude errors in the assignment of interpreters to speeches, as even careful inspection of a speech sometimes does not reveal the identity of an interpreter.

In order to be able to compare the German translations and interpretations with original (non-translated) German data, we moreover created two further reference corpora. First, we extracted a random sample of 50 speeches (verbatim reports) delivered in German from the Europarl-direct corpus (Koehn 2005; Cartoni and Meyer 2012). Second, we created a small corpus of political interviews in German, as previous studies have prominently focused on the register properties of interpreted speech (e.g., Shlesinger 2008; Shlesinger and Ordan 2012). We refer to these transcripts as the “interview corpus” in the following. It is intended to serve as a reference point for unscripted political speech. The German TV channel ZDF conducts regular 20 min summer interviews with prominent politicians. We used five of these interviews, conducted in 2022, for our corpus.⁶ We obtained the automatically generated subtitles from YouTube and corrected them manually using Hexatomic. The corpus contains 14,721 tokens including punctuation (Lang: 3507 words; Lindner: 2795 words; Merz: 3005 words; Scholz: 2680 words; Weidel: 2734 words).

All texts—the English and German verbatim reports (translations as well as originals), the transcripts of the German interpretations and the interviews—were processed in the same way. Initial address formulas (“Mister President”) were removed, as they were uninformative and introduced a bias, specifically when comparing the Europarl material to other types of corpus material, such as interviews. The texts were annotated with the Python package `stanza`, a state-of-the-art neural-network-based parser.⁷ We used the tagset of the Universal Dependencies (UD) framework⁸ for part-of-speech annotation.⁹ The UD-tags are designed for crosslinguistic comparability, and are thus relatively coarse-grained, distinguishing only 17 classes (see Appendix A.2).¹⁰ Given that translations and interpretations were created differently (manual vs. automatic transcription), some tags were not used for the analysis as they were not (reliably) recorded in at least one text type, specifically INTJ (interjection), PUNCT (punctuation) and SYM (symbol). Moreover, the tag X (for “other”) was disregarded.

Given that the tagger did not tag proper names (PROPN) reliably in both languages, we subsumed proper names under nouns. For example, the tagger categorized the English word *Presidency* as a proper name, probably because it is capitalized (e.g., in *Swedish Presidency*), while the German tagger categorizes the corresponding noun *Ratsvorsitz* as NOUN.

A major difficulty in comparing English and German concerns the fact that English has many compounds whose components are separated by a white space in writing. Accordingly, counts of English nouns tend to be inflated in comparison to German, where compounds are spelled in one word. The `stanza`-parser identifies (modifying) elements of compounds as such and assigns a (dependency) tag to them. For example, *climate change* in *tackling climate change* is parsed as `climate_NOUN_compound change_NOUN_obj`. In order to keep the numbers between English and German comparable, compounds of this type were only counted as one noun. (Technically, the `NOUN_compound`-tags were simply removed from the counts.)

Following a suggestion from an anonymous reviewer, we systematically split up two tags, as they neutralized potentially informative categories: we distinguished between attributive vs. predicative adjectives and between the third-person singular neuter pronouns *it* and *es* vs. other personal pronouns. The distinction between attributive and predicative

adjectives could be made because the stanza-parser returns dependency tags, and there is a tag “amod” that specifically identifies attributive adjectives (represented as ADJ.amod in the following, as opposed to ADJ.pred). The pronouns *it* and *es* are distinguished from other pronouns by a tag of their own (PRONIT vs. PRON, the latter standing for pronouns other than *it/es*). Moreover, we split the PART-tag into the specific words that it generalizes over. In English, PART covers three elements, the Saxon genitive marker ‘s, the negation particle *not* and the infinitive marker *to*. In German, it only covers the negation marker *nicht* and the infinitive marker *zu*. As there is no German element that would be comparable to the Saxon genitive ‘s, since genitives are realized as (often unsegmentable) inflection, we disregarded the Saxon genitives in the analysis and split up PART into two tags, NEGPART (covering Engl. *not* and Germ. *nicht*) and INFPART (covering the infinitive markers *to* and *zu*). INFPART was not used for the quantitative analysis, however, because it delivers misleading results. English *to* is naturally more frequent than German *zu* (as an independent word), as in the latter language the infinitive markers are often morphologically integrated into the verb. This happens with verbs carrying a prefix, like *umfallen* ‘fall over’, whose infinitive is *umzufallen*.

Note that all analyses were run for both tagsets, the “basic” one (as delivered by the tagger) and the “extended” one, with the additional differentiations described above.

3. Results

3.1. English Originals, German Translations and German Interpretations

In a first step, we compared the frequencies of the part-of-speech tags under analysis in original speeches, in the written translations and in the interpretations. The pairwise comparisons of twelve tags are shown in Figure 1 (trl: translation, interp: interpretation), with the results of a Wilcoxon signed-rank test (with matched pairs). The figures show the data for the extended tagset, with differentiations between the types of adjectives and pronouns. The corresponding plots for the comparisons between the originals and translations, and the originals and interpretations, are provided in Figures A3 and A4 in Appendixes A.5 and A.6. The plots for the basic tagset are provided in the Supplementary Materials.

In order to understand the interplay of the various POS-tags, we ran a Correspondence Analysis (CA). The CA for the translated and interpreted texts shows a clear separation between nominal elements, on the one hand, and verbal or clause-level elements, on the other, along the first dimension (accounting for 55.6% of the variance with the basic tagset and 51.9% of the variance with the extended tagset). Figure 2 shows the column scores (corresponding to variables rather than texts) of the tags for Dimension 1 (for both tagsets). The tags in the negative range are nominal; those in the positive range are located at the clause level. It thus seems reasonable to regard Dimension 1 of the Correspondence Analysis as a measure locating texts on a scale of a more nominal to a less nominal style.

The Dimension 1 scores of the CA fitted to the translation and interpreting data differ significantly between the two text types “translation” and “interpretation” ($p < 0.001$, according to a Wilcoxon signed-rank test). Figure 3 shows the pairwise comparisons of the texts in the sample. The picture is remarkably clear: all the interpreted texts exhibit a less nominal style than the corresponding translations, if the Dimension 1 scores of the CA are regarded as operationalizations of a nominal style.

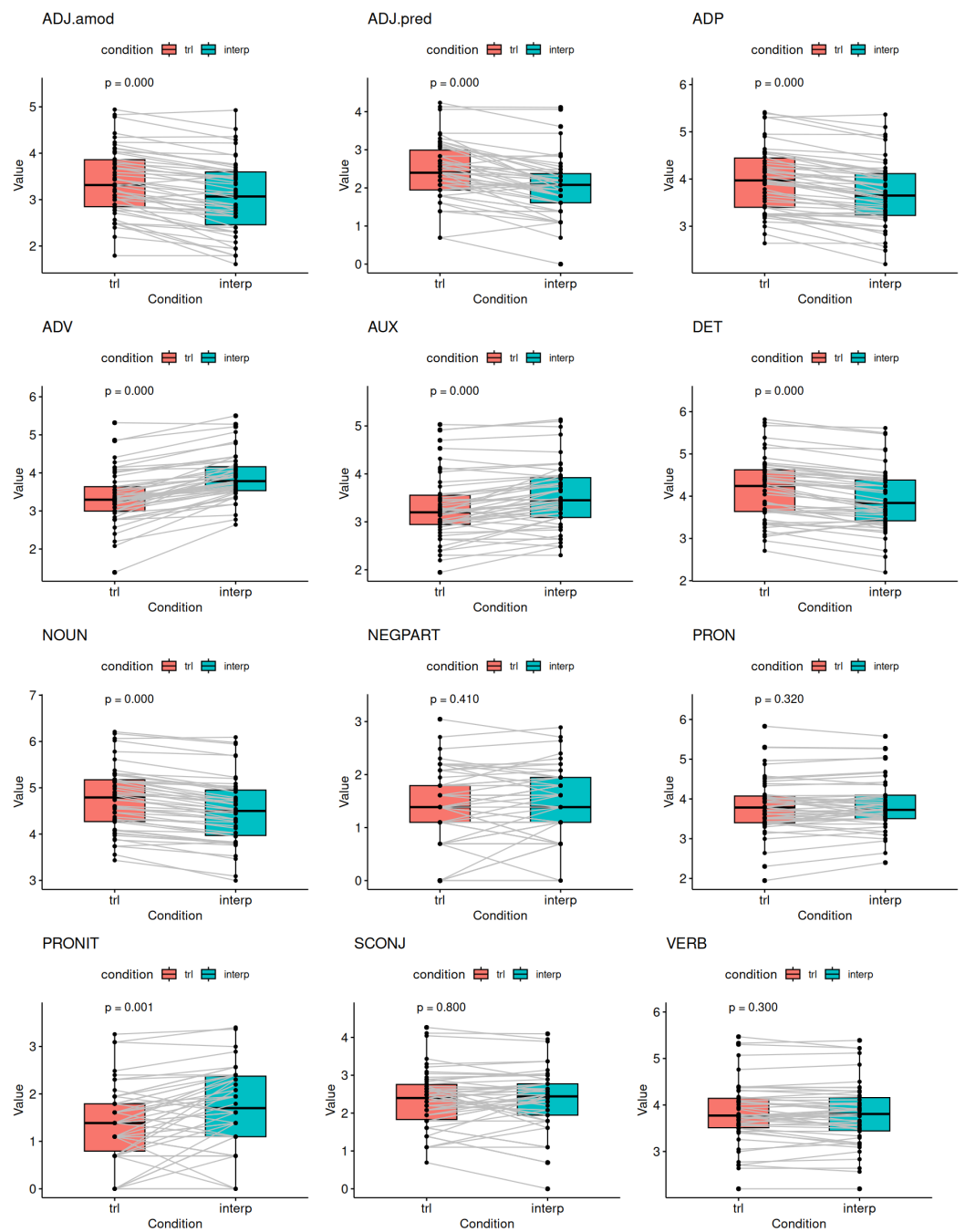


Figure 1. Pairwise comparisons of POS-frequencies (log-transformed) per speech in translations and interpretations. Each line connects data points representing the same speech (trl: translation, interp: interpretation).

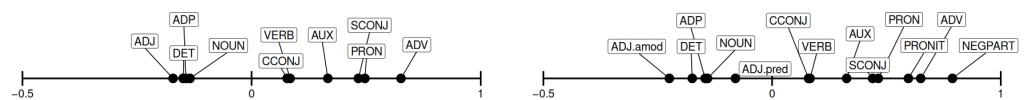


Figure 2. Column scores (variables) for Dimension 1 in a Correspondence Analysis of the POS-tag frequencies in translations and interpretations (left: basic tagset, right: extended tagset).

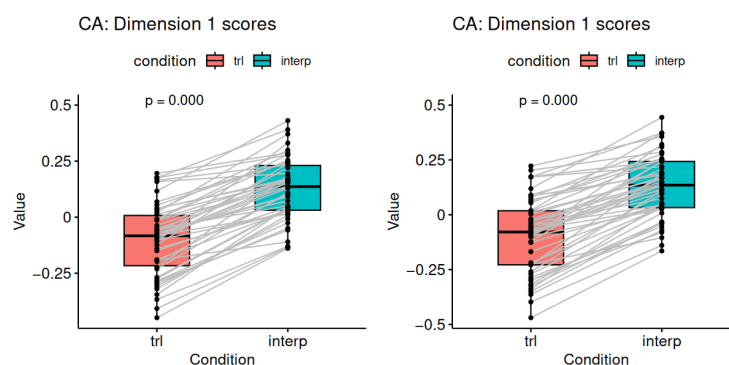


Figure 3. Comparison of row scores (texts) for Dimension 1 in a Correspondence Analysis of the POS-tag frequencies in translations and interpretations (left: basic tagset, right: extended tagset).

In order to obtain precise, multivariate statistical measures of significance and confidence, we fitted a mixed-effects Poisson regression model to the data, with a random intercept for the individual speeches. A random effect for “interpreter” did not prove informative. The model is visualized in Figure 4. The plot at the top shows for each tag (from the extended tagset) the predicted counts, adjusted to the average text length (the plots for the basic tagset are provided in the Supplementary Materials). The plot at the bottom visualizes the post hoc comparisons between the originals and interpretations as well as the translations (the p -values for the post hoc comparisons are provided in Figure A5 in Appendix A.6). The y-axis shows the ratio of the number of tags in the originals to the number of tags in the interpretations (red) and translations (blue), again with 95% confidence intervals for the estimates. A value of 1 indicates that the number of tags in the original and the translation (red) or interpretation (blue) are identical. A ratio of >1 indicates that the tag is (relatively) more frequent in the originals. Note that the y-axes were log-transformed in both plots.

To illustrate how the two plots in Figure 4 are interpreted, let us consider the NOUN-tag (remember that the values on the y-axis are log-transformed, which “distorts” the visual impression). The predicted frequencies for the three text types are 111.07 (orig), 90.46 (interp) and 114.96 (trl) for a hypothetical text of 564.54 words. The regression model estimates that the nouns in the originals will be 1.23 times more frequent than in the interpretations. This is where the red point in the bottom plot is located (the y-axis is also log-transformed), and it is also the ratio of 111.07 (originals) to 90.46 (interpretations).

Let us now consider the frequencies of the tags, starting with the parts of speech that are over-represented in the interpretations, in comparison to both the originals and translations. These are the tags whose 95% confidence intervals are located entirely underneath the line, indicating a 1/1 ratio in the bottom plot, and which are significantly lower than the estimates for the translations (see Figure A5 for the exact p -values). This condition is met by two tags, ADV and PRONIT. The low position of the ADV-tag in the bottom plot shows that the ratio of the counts in the originals and interpretations is very low—it is ≈ 0.51 , which means that adverbs are almost twice as frequent in the interpretations than in the originals. The ratio for the PRONIT-tags is ≈ 0.82 .

A second group of tags is over-represented in the interpretations relative to the originals but not relative to the written translations. This group contains NEGPART (the word *nicht*) and PRON (pronouns other than *es*).

One tag is over-represented in the interpretations relative to the written translations but not relative to the originals: AUX, covering the copulas (Engl. *be* and Germ. *sein*) and modals such as Engl. *will*, *can* and *may* and Germ. *werden*, *können*, *müssen*, etc.

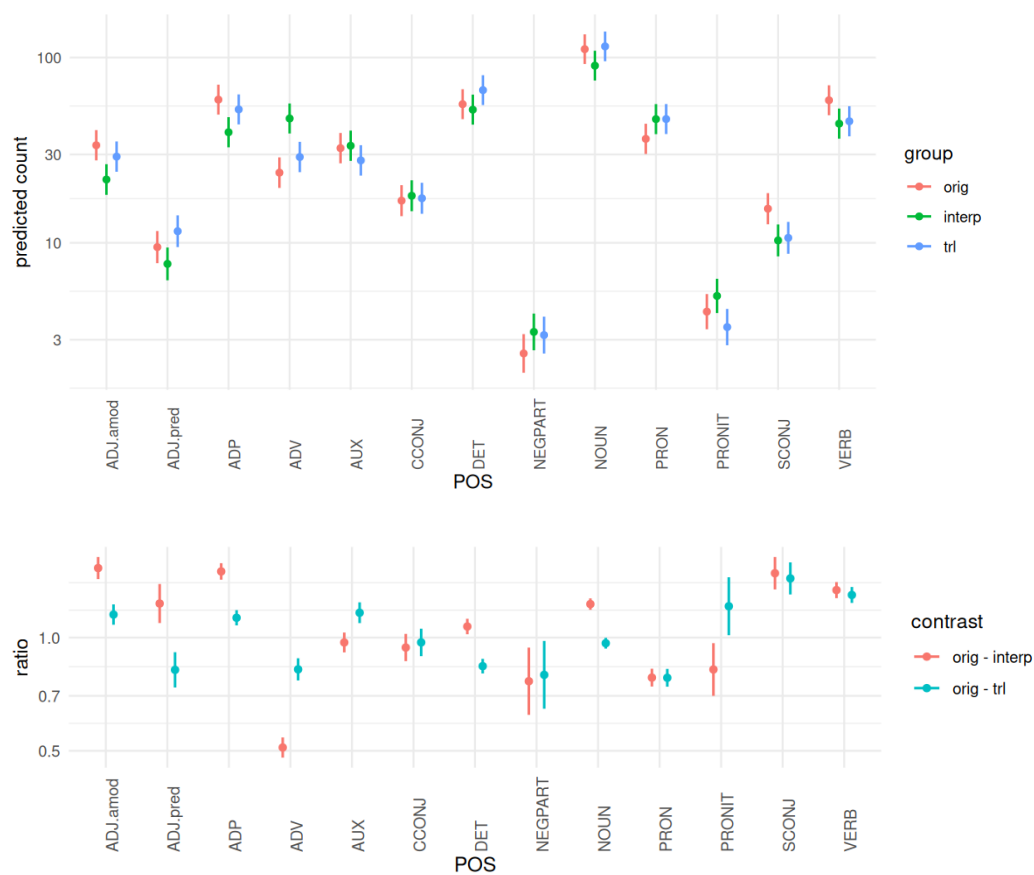


Figure 4. Top: predicted frequencies (log-transformed) for text of average length (564.5 words), with 95% confidence intervals; bottom: pairwise post hoc comparisons, indicating the ratios of tags, with 95% confidence intervals.

The following tags are underrepresented in the interpretations relative to both the originals and translations: ADJ.pred (predicative adjectives), ADJ.amod (attributive adjectives), ADP (adpositions), DET (determiners) and NOUN (nouns and proper names).

Finally, there is a group of tags that are underrepresented in the interpretations relative to the originals but not relative to the translations: SCONC (subordinating conjunctions) and VERB (main verbs).

The (quantitatively) most striking result is the rarity of the nominal elements in the interpretations. Not only nouns but also determiners, attributive adjectives and adpositions are significantly underrepresented in the interpretations. There seems to be a general tendency toward the avoidance of nouns in interpretations.

Another interesting result is the high frequency of adverbs and pronouns in the interpretations. This matches the results of Shlesinger and Ordan (2012) for English–Hebrew interpretations, where pronouns and adverbs also stand out as prominent characteristics of interpretations (the results of Dayter 2018 are ambiguous with respect to pronouns; she finds a high proportion of pronouns in Russian interpretations of English texts but not in English interpretations of Russian texts).

3.2. English Originals, German Mediated Language and German Originals

Having compared the two mediated text types—German translations and interpretations—with English originals, we can now bring German originals into the picture. We ran a Correspondence Analysis with data from all the text types, including the German speeches and the political interviews. Consider, first, the scores for the columns (variables) in Dimensions 1 and 2, shown in Figure 5 on the left (based on the extended tagset). As in the CA based on the translated and interpreted data only, Dimension 1 clearly separates the

nominal from the non-verbal/clause-level elements. Dimension 1 accounts for 57.6% of the variance. Dimension 2 is less informative, accounting for 11% of the variance. Positive values on Dimension 2 are primarily driven by subordinating conjunctions (positive values) and adverbs (negative values). This shows that in the data used for our study, there is a certain complementarity between complex sentences and adverbs.

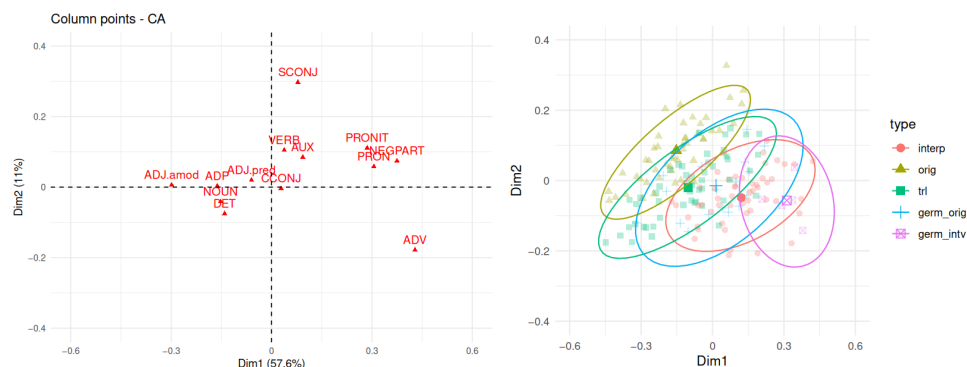


Figure 5. Correspondence Analysis based on all text types (originals, translations and interpretations), columns (variables) at the left and rows (texts) at the right.

The plot on the right-hand side of Figure 5 shows the distributions of the texts relative to the two-dimensional space constituted by Dimensions 1 and 2 (with the centroids and ellipses indicating 95% confidence intervals). The English originals and the German translations exhibit the lowest scores on this dimension and are thus the most nominal text types. The interviews are located at the right end of the plot, exhibiting the least nominal style. The translations and interpretations are located in-between.

Against the background of the observations made by Shlesinger and Ordan (2012), it is of interest whether the interpretations are located closer to the German originals or to the interviews. A linear model shows that, in fact, the German interpretations differ from both of these text types along Dimension 1 ($p = 0.001$ for the originals, and $p < 0.001$ for the interviews) as well as Dimension 2 ($p = 0.004$ for the originals and $p = 0.001$ for the interviews.).

We can now compare the five text types under analysis—English originals, German originals, German written translations, German interpretations and German interviews—by considering the results of a Poisson regression model. Figure 6 shows the predicted counts with 95% confidence intervals.

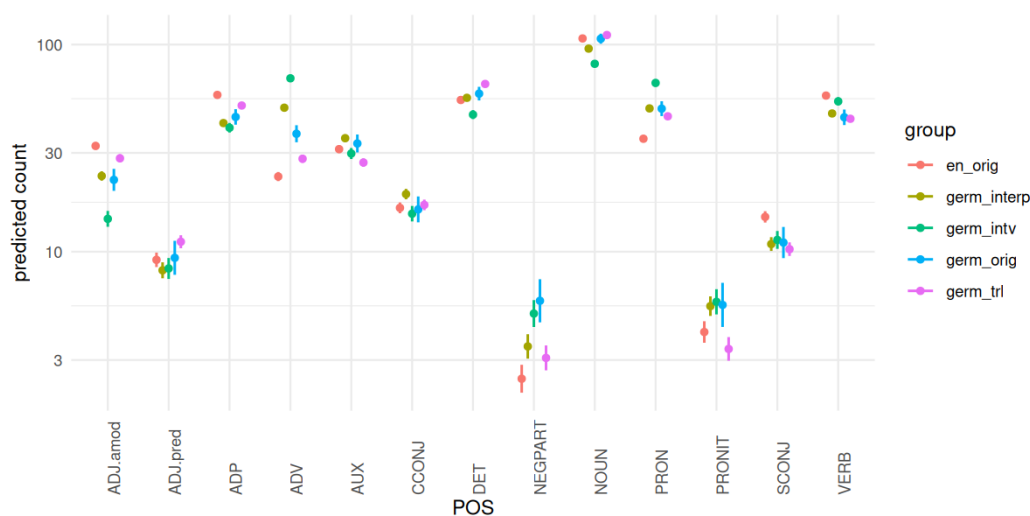


Figure 6. Predicted frequencies (log-transformed) for text of average length, with 95% confidence intervals, according to a Poisson model.

Let us reconsider the findings of the comparison between the English originals, German translations and German interpretations. Adverbs and third-person neuter pronouns were found to be typical of the interpretations. Adverbs are also prominent in the German interviews and can probably be regarded as a feature of spoken language. The third-person neuter pronoun *es* is also widely used in the German original speeches (in the Europarl corpus), as well as in the interviews. It is rare in the German translations, however. Perhaps the translations adapt to the English originals here (*shining through*), which also exhibit a relatively low number of *its*.

Auxiliaries were shown to be more frequent in the interpretations than in the translations, though they were also frequent in the English originals. As Figure 6 shows, auxiliaries are relatively common in the original German speeches, too.

Negative particles and pronouns other than *es* were found to be over-represented in the translations and interpretations. Figure 6 shows that negative particles are particularly frequent in the German originals and interviews. In comparison with the original data, they are actually underrepresented. Similarly, pronouns other than *es* are also frequent in the interviews and the German Europarl speeches.

Elements of the nominal structure were shown to be underrepresented in the interpretations. Figure 6 shows that this is typical of the spoken medium, as the interviews also feature low frequencies of adjectives, nouns, determiners and adpositions. In fact, the estimated counts for the interviews are lower for all the nominal categories than those for the interpretations.

Finally, the main verbs and subordinating conjunctions were shown to be infrequent in the translations as well as the interpretations, in comparison to the English originals. Figure 6 shows that the counts for the German originals are also low, confirming the common assumptions about a general tendency for German to lean toward a nominal style, in comparison with English. Even German interviews, the “most verbal” text type, have lower counts for verbs than the English originals. The same goes for subordinating conjunctions: they are more frequent in English than in any of the German text types.

4. Discussion and Conclusions

A recurrent theme in the comparison of the translated and interpreted language is the relative distribution of nouns and verbs. The present study has confirmed previous findings showing that there are asymmetries between the translations and interpretations in the distributions of nouns and verbs. However, it was pointed out that the relationship between nouns and verbs is not symmetric in the data used for the present study. While the interpretations exhibit a low number of nominal material, and a high number of pronouns, they do not feature a particularly high number of verbs.

A closer inspection of the data can shed more light on the distributions of the parts of speech in the interpreting and translations. First, there is a clear tendency for interpreters to “economize” on nouns. Conversely, as we saw, the interpretations contain more pronouns than the originals and translations in our data. In (4), the speaker talks about *small or large countries* (*kleine ebenso wie große Länder* in the translation, cf. (5)). The interpretation simply has *jeder* ‘everybody’ (cf. (6)). This example can probably be regarded as a case of implicitation.

- (4) ...government leaders who appear only to be interested in sharing out jobs amongst themselves, be it from small or large countries ...
- (5) ...Staatsoberhäuptern, die nur daran interessiert scheinen, sich gegenseitig Jobs zuzuschieben. Dies betrifft kleine ebenso wie große Länder ...
- (6) ...jetzt wird hier eine Art Feilscherei zwischen verschiedenen Mitgliedstaaten betrieben, jeder meldet seine Interessen an ...

As was seen, the interpretations exhibit a high number of adverbs in comparison to the translations. These adverbs are often deictic, and they can be used—unlike in the written translations—because the speakers and addressees are situated in a spatio-temporal environment to which interpreters can refer. In (9), the deictic adverbs *jetzt* ‘now’ and

hier ‘here’ are used. Note that the original in (7) contains no such adverbs nor does the translation in (8).

- (7) Instead, the discussion is degenerating into a shabby debate between government leaders ...
- (8) Stattdessen entwickelt sich die Diskussion in eine schäbige Debatte zwischen Staatsoberhäuptern ...
- (9) ... aber jetzt wird hier eine Art Feilscherei zwischen verschiedenen Mitgliedstaaten betrieben.

There is a close relationship between the use of pronouns and specific adverbs. Both types of expressions require situational accessibility. In some cases—probably even in (9)—deictic adverbs have a quasi-anaphoric function, referring back to a topic. The deictic adverb *hier* ‘here’ in (12) refers back to *die Lösung des Klimawandels* ‘the solution to climate change’, and it could be paraphrased with ‘in this matter’, which is also the wording found in the translation (*in dieser Angelegenheit*, cf. (11); in English, *the kind of issue*, cf. (10)).

- (10) Tackling climate change is one of our highest priorities and is the kind of issue where we expect and want the European Union to take a strong lead.
- (11) In dieser Angelegenheit erwarten und wollen wir, dass die Europäische Union eine starke Führungsrolle übernimmt.
- (12) ... die Lösung des Klimawandels ist für uns die höchste Priorität und wir erwarten von der EU hier eine starke Führungsrolle.

We saw that auxiliaries were comparatively frequent in the interpretations. As mentioned above, the models used for the annotation of the data classify copulas and modals as AUX, in English as well as German. The higher incidence of modals in the interpretations sometimes seems to be a result of syntactic restructuring. In (13) (from speech 20080903_21063000), *ought to* expresses a deontic modality that (implicitly) takes scope over the entire sentence. All predications contained in the sentence are thus modalized: (i) *do that in a non-intrusive way*, (ii) *that preserves the right of third countries ...* and (iii) *while opening them to external supply if competition is favoured* all describe desirable states of affairs:

- (13) We ought to do that in a non-intrusive way that preserves the right of third countries to regulate the different services sectors domestically as they wish, while opening them to external supply if competition is favoured.

The translation of (13) uses a coordinating structure, the second part of the conjunct being modified by a finite subordinate clause. In this way, modalization is also preserved for the entire sentence:

- (14) Hier sollten wir behutsam vorgehen und das Recht von Drittländern achten, ihre unterschiedlichen einheimischen Dienstleistungssektoren so zu regulieren, wie sie es wünschen, während sie zugleich für ausländische Angebote erschlossen werden, wenn Wettbewerb vorgezogen wird.

The interpreter chose to split up the sentence into two main clauses, the second one being modified by a finite adverbial clause. The modality therefore needs to be expressed in each clause separately:

- (15) Wir dürfen hier nicht zu dirigistisch vorgehen. Wir müssen die Rechte von Drittländern respektieren, verschiedene Dienstleistungssektoren national nach ihren eigenen Vorstellungen zu regeln, wobei es natürlich eine Förderung der externen Möglichkeiten geben muss.

As the discussion has shown, it is mostly possible to reconstruct a “motivation” for a specific choice made by a translator or interpreter. To what extent these motivations are psychologically real is a different question. In Section 1, three main types of explanation were mentioned for choices made by translators or interpreters: processing constraints, the factor of register and interpreting strategies. Considering the examples above, it is hard to imagine how these factors could be kept apart. Consider the use of deictic adverbs. The

spatio-temporal context is a prerequisite (necessary condition) for using words such as *here*. But what exactly triggers the use of an adverb of this type? Is it a strategic choice, a cognitive routine giving preference to deictic expressions when they are available—or perhaps a strategic choice that has turned into a cognitive routine? Or is it automatic behavior avoiding complexity? It is hard to think of an experimental setup that could help us answer such questions.

A more feasible endeavor seems to be the project of understanding the relationship between mediation and mode. Does interpreting relate to written translation in the same way as speaking relates to writing? Are the effects of mediation and communicative mode additive or perhaps multiplicative? Or is interpreting, as a cognitive process, fundamentally different from written translation? Addressing such questions will obviously require both observational and experimental studies and appropriate methodologies. The present study is intended as a modest contribution to this endeavor, which, on the observational side, implies corpus building—including the automatic processing of large amounts of data, as described in Section 2, and appropriate methodologies for quantitative analyses, as applied in Section 3.

Supplementary Materials: The Supplementary Material with the data and scripts can be downloaded at the following OSF-repository <https://osf.io/tp86c/> (accessed on 27 January 2023). They contain the transcriptions (raw and tagged); the data frames with POS-frequencies, with and without the PART-split in the German data; the data frames for the identification of interpreters; and an R-script for the statistical analysis and generation of plots as well as the plots.

Author Contributions: Conceptualization, both authors; methodology, both authors; software, R.B.; validation, both authors; resources, both authors; data curation, both authors; writing—original draft preparation, V.G.; writing—review and editing, both authors; visualization, V.G.; funding acquisition, V.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG) grant number 391160252.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and scripts are made available in the Supplementary Materials (<https://osf.io/tp86c/>, accessed on 17 January 2023).

Acknowledgments: We wish to thank the audiences at the conferences ‘Word Formation Theories VI & Typology and Universals in Word Formation V’ (Košice, 23–26 June 2023) and ‘Translation in Transition’ (Prague, 22–23 September 2022) for valuable feedback.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Speeches and Speakers

Table A1. Speeches and speakers.

EAF-File	Europarl-File	Speaker-Info	Nationality	Length (ms)
20091111_16533400_16560100	ep-09-11-11-013	Timothy Kirkhope	UK	122,500
20080903_21063000_21130900	ep-08-09-03-016	Peter Mandelson	UK	318,220
20080901_23043500_23093900	ep-08-09-01-023	Nickolay Mladenov	Bulgaria	281,660
20090422_09103100_09153900	ep-09-04-22-004	Peter Skinner	UK	102,621
20091111_16135700_16240100	ep-09-11-11-013	Fredrik Reinfeldt	Sweden	565,480
20090916_11103100_11162100	ep-09-09-16-006	Paweł Samecki	Poland	191,040
20080903_22090600_22115100	ep-08-09-03-017	Robert Evans	UK	106,520
20100407_15323600_15362400	ep-10-04-07-004	Timothy Kirkhope	UK	179,460
20090114_23070200_23105200	ep-09-01-14-014	Richard Howitt	UK	164,260
20090915_19312000_19354200	ep-09-09-15-012	Karel De Gucht	Belgium	175,640
20091007_15411400_15474800	ep-09-10-07-016	Fredrik Reinfeldt	Sweden	363,380
20080904_16234700_16280600	ep-08-09-04-012-03	Benita Ferrero-Waldner	Austria	240,060
20091216_16542500_16595400	ep-09-12-16-010	Karel De Gucht	Belgium	118,230
20100224_16084800_16113300	ep-10-02-24-013	Timothy Kirkhope	UK	130,610
20100224_15525600_15582200	ep-10-02-24-013	Stephen Hughes	UK	302,390
20080901_19114600_19141400	ep-08-09-01-018	Urszula Gacek	Poland	142,540
20080903_21560700_22004700	ep-08-09-03-017	Peter Mandelson	UK	241,869
20080902_15474800_16061000	ep-08-09-02-010	Viviane Reding	Luxemburg	978,704
20080903_18263700_18301500	ep-08-09-03-015	Meglana Kuneva	Bulgaria	19,0130
20080901_23355400_23413900	ep-08-09-01-024	László Kovács	Hungary	320,340
20080903_22372500_22395600	ep-08-09-03-017	Timothy Kirkhope	UK	111,290
20091217_09505600_09580300	ep-09-12-17-003	Paweł Samecki	Poland	388,840
20090218_21480900_21503100	ep-09-02-18-023	Toomas Savi	Estonia	107,100
20090310_21084000_21105900	ep-09-03-10-017	Philip Bushill-Matthews	UK	71,740
20100211_10080200_10103400	ep-10-02-11-004	Timothy Kirkhope	UK	42,650
20100407_14573200_15113200	ep-10-04-07-004	Maroš Šefčovič	Slovakia	802,670
20080902_22413600_22481300	ep-08-09-02-016	Viviane Reding	Luxemburg	331,758
20091126_15414900_15463100	ep-09-11-26-012-02	Karel De Gucht	Belgium	42,460
20080903_18385000_18410200	ep-08-09-03-015	Charlie McCreevy	Ireland	113,290
20080903_19021500_19032400	ep-08-09-03-015	Robert Evans	UK	68,630
20080903_21490100_21535100	ep-08-09-03-016	Peter Mandelson	UK	267,619
20090715_15154400_15182700	ep-09-07-15-008	Richard Howitt	UK	63,910
20090914_22243100_22302200	ep-09-09-14-023	Karel De Gucht	Belgium	99,800
20080902_23445200_23465800	ep-08-09-02-017	Androula Vassiliou	Greece	90,200
20090421_23283400_23312300	ep-09-04-21-027	Joe Borg	Malta	58,570
20090915_19233400_19282900	ep-09-09-15-012	Karel De Gucht	Belgium	154,720
20100325_11283700_11313400	ep-10-03-25-003	Edward Scicluna	Malta	157,090
20080903_21264100_21295100	ep-08-09-03-016	Caroline Lucas	UK	16,1440
20090218_16040100_16080500	ep-09-02-18-014	Kathy Sinnott	Ireland	211,030
20091217_09431100_09453600	ep-09-12-17-002	Paweł Samecki	Poland	101,441
20090219_12451800_12465300	ep-09-02-19-010	Richard Corbett	UK	65,540
20080904_11155500_11175000	ep-08-09-04-005	Miroslav Ouzký	Czech Republic	86,560
20080904_12011000_12040900	ep-08-09-04-005	Stavros Dimas	Greece	158,680
20100310_11363400_11481900	ep-10-03-10-006	Catherine Ashton	UK	67,180
20080902_21063800_21111100	ep-08-09-02-013	Proinsias De Rossa	Ireland	223,660
20090402_10332800_10353100	ep-09-04-02-006	Linda McAvan	UK	81,660
20090914_22464400_22545300	ep-09-09-14-023	Karel De Gucht	Belgium	231,000
20090915_19040500_19075100	ep-09-09-15-012	Stavros Dimas	Greece	139,950
20090312_15265900_15281500	ep-09-03-12-013-02	Marios Matsakis	Greece	33,740
20090311_15554000_16114600	ep-09-03-11-012	Alexandr Vondra	Czech Republic	57,669

Appendix A.2. Universal Dependencies Tagset

Table A2. Universal dependencies tagset (see <https://universaldependencies.org/u/pos/index.html>, accessed on 27 January 2023).

ADJ:	adjective
ADP:	adposition
ADV:	adverb
AUX:	auxiliary
CCONJ:	coordinating conjunction
DET:	determiner
INTJ:	interjection
NOUN:	noun
NUM:	numeral
PART:	particle
PRON:	pronoun
PROPN:	proper noun
PUNCT:	punctuation
SCONJ:	subordinating conjunction
SYM:	symbol
VERB:	verb
X:	other

Appendix A.3. Density Plot for Sample

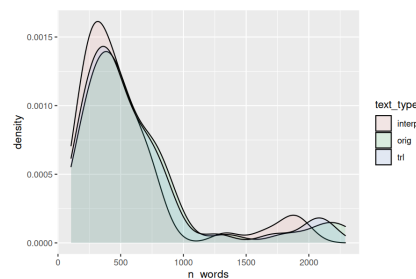


Figure A1. Density plots for the number of words per speech in the sample, for the three text types “original”, “translation” and “interpretation”. The x-axis shows the number of words, the y-axis the probability density for the values on the x-axis.

Appendix A.4. Dimension 1 Scores for Interpreters and Translators

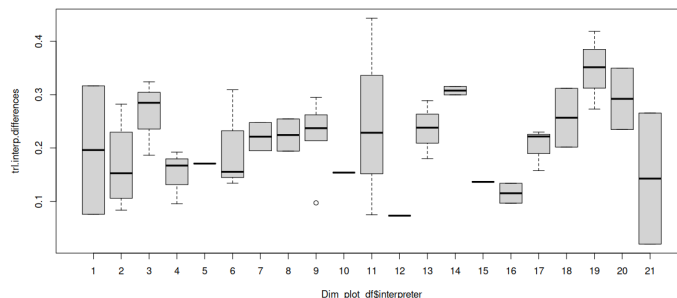


Figure A2. Differences between Dimension 1 scores for interpretations and translations per interpreter.

Appendix A.5. Pairwise Comparisons: Originals and Translations

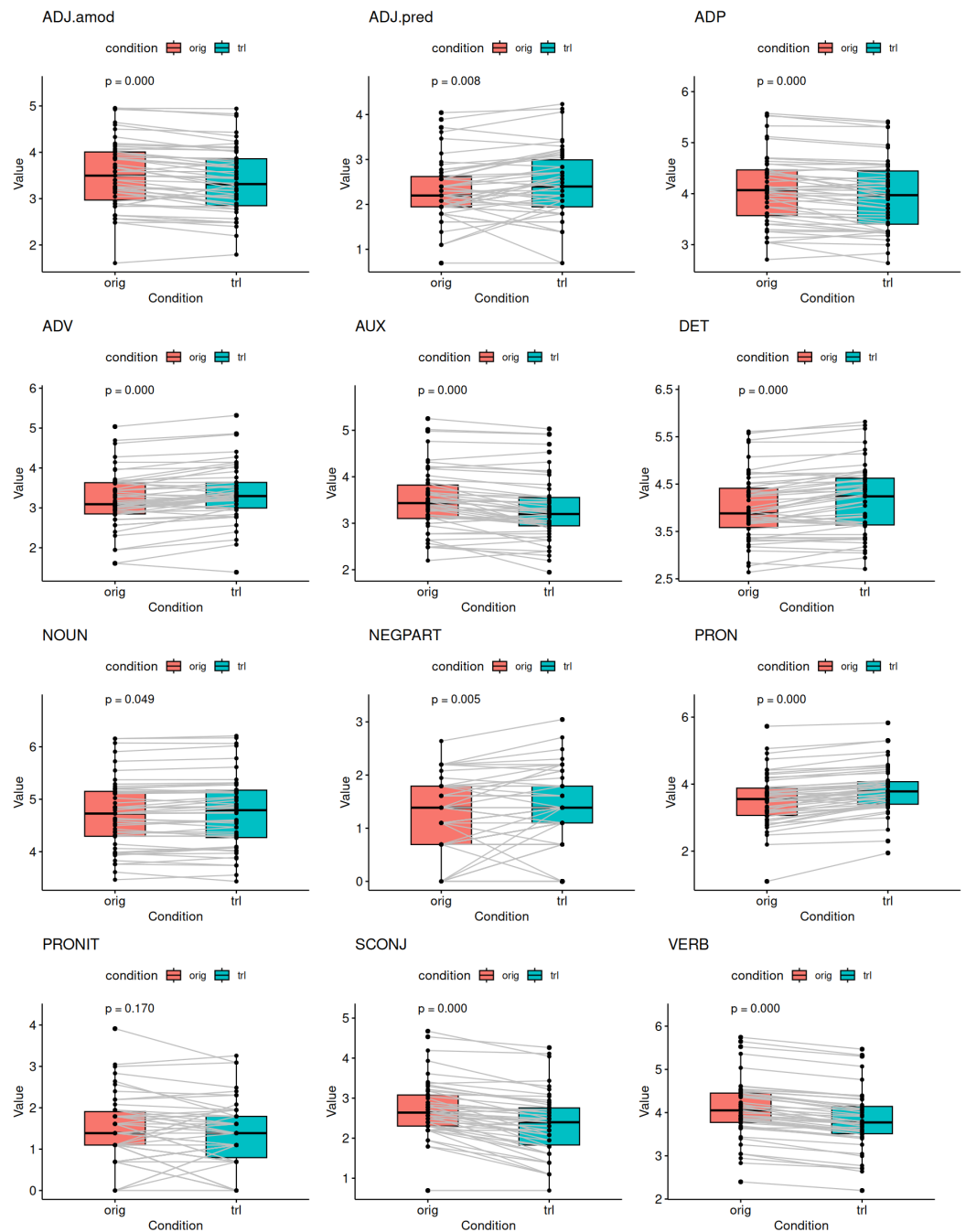


Figure A3. Pairwise comparisons of POS-frequencies (log-transformed) per speech in originals and translations. Each line connects data points representing the same speech (orig: original, trl: translation).

Appendix A.6. Pairwise Comparisons: Originals and Interpretations

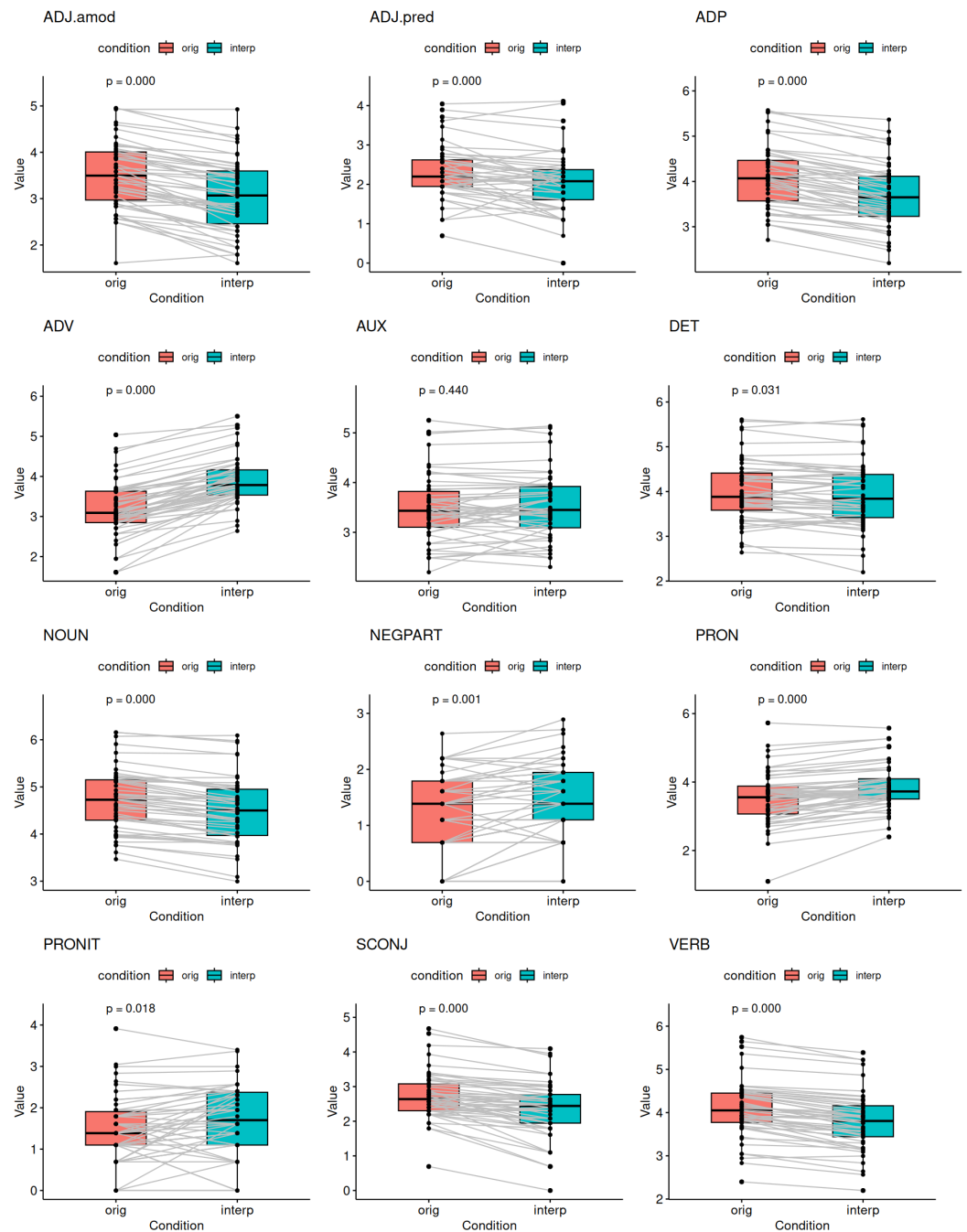


Figure A4. Pairwise comparisons of POS-frequencies (log-transformed) per speech in originals and translations. Each line connects data points representing the same speech (orig: original, interp: interpretation).

Appendix A.7. *p*-Values for Post Hoc Comparisons

Figure A5. *p*-values for post hoc comparisons (top: basic tagset, bottom: extended tagset).

Notes

- ¹ Note also that *the people needed to fill them* disappeared in the interpretation. This, too, can be regarded as implicature, and it illustrates how the availability of syntactic structures can determine translation choices. German does not have nonfinite postnominal relative clauses of the type *needed to fill them*. Moreover, the German correlate of the verb *fill* would be awkward here. Interestingly, even the translation in (1) does not render this part of the sentence.
- ² The files are property of the European Union represented by the European Parliament; see <https://www.europarl.europa.eu/legal-notice/en/> (accessed on 27 January 2023) for the conditions of use.
- ³ <https://hexatomic.github.io/> (accessed on 27 January 2023)
- ⁴ The speeches 20080903_2149, 20080903_2237, 20090219_1245 and 20090422_0910 were delivered extemporaneously. The speakers of speeches 20090715_1515, 20091007_1541 and 20091111_1613 speak extemporaneously but have a script.
- ⁵ <https://github.com/resemble-ai/Resemblyzer> (accessed on 27 January 2023)
- ⁶ The interviews were conducted with:
O. Scholz (<https://www.youtube.com/watch?v=Xs2LuGq0eNU>),
R. Lang (<https://www.youtube.com/watch?v=wHk3xr3XLVg>),
C. Lindner (<https://www.youtube.com/watch?v=IOoVjwvobqE>),
A. Weidel (https://www.youtube.com/watch?v=VpQobk_ybwc), and
F. Merz (<https://www.youtube.com/watch?v=XhoOH6bJDeU>),
all accessed on 10 January 2023.
- ⁷ <https://stanfordnlp.github.io/stanza/> (accessed on 27 January 2023)
- ⁸ <https://universaldependencies.org/> (accessed on 27 January 2023)
- ⁹ The models for English and German are described at https://universaldependencies.org/treebanks/en_ewt/index.html and https://universaldependencies.org/treebanks/de_gsd/index.html (both accessed on 27 January 2023).
- ¹⁰ The tagset can also be inspected at <https://universaldependencies.org/u/pos/index.html> (accessed on 27 January 2023).

References

- Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In *Text and Technology: In Honour of John Sinclair*. Edited by Mona Baker, Gill Francis and Elena Tognini-Bonelli. Amsterdam and Philadelphia: Benjamins, pp. 233–50.
- Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Edited by Harold Somers. Amsterdam: John Benjamins, pp. 175–86.

- Bendazzoli, Claudio, and Annalisa Sandrelli. 2005. An approach to corpus-based interpreting studies: Developing EPIC (European Parliament Interpreting Corpus). In *Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation*. Edited by Heidrun Gerzymisch-Arbogast and Sandra Nauert. Saarbrücken: Advanced Translation Research Center, pp. 149–60.
- Bernardini, Silvia, Adriano Ferraresi, Mariachiara Russo, Camille Collard, and Bart Defrancq. 2018. Building interpreting and intermodal corpora: A how-to for a formidable task. In *Making Way in Corpus-Based Interpreting Studies*. Edited by Mariachiara Russo, Claudio Bendazzoli and Bart Defrancq. Singapore: Springer, vol. 1, pp. 21–42.
- Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Cartoni, Bruno, and Thomas Meyer. 2012. Extracting directional and comparable corpora from a multilingual corpus for translation studies. Paper presented at Proceedings 8th International Conference on Language Resources and Evaluation (LREC), Istanbul, Turkey, May 21–27.
- Chmiel, Agnieszka, Danijel Korzinek, Marta Kajzer-Wietrzny, Przemysław Janikowski, Dariusz Jakubowski, and Dominika Polakowska. 2022. Fluency parameters in the polish interpreting corpus (PINC). In *Empirical Investigations into the Forms of Mediated Discourse at the European Parliament*. Edited by Marta Kajzer-Wietrzny, Adriano Ferraresi, Ilmari Ivaska and Silvia Bernardini. Berlin: Language Science Press, pp. 63–91.
- Christoffels, Ingrid K., and Annette M. B. de Groot. 2005. Simultaneous interpreting: A cognitive perspective. In *Handbook of Bilingualism: Psycholinguistic Approaches*. Edited by Judith F. Kroll and Annette M. B. de Groot. Oxford: Oxford University Press, pp. 454–79.
- Dayter, Daria. 2018. Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN). *FORUM* 16: 142–64. [\[CrossRef\]](#)
- Dayter, Daria. 2021. Strategies in a corpus of simultaneous interpreting. *Meta: Translators' Journal* 66: 594–617. [\[CrossRef\]](#)
- De Groot, Annette M. B. 1997. The cognitive study of translation and interpretation: Three approaches. In *Cognitive Processes in Translation and Interpreting*. Edited by Joseph H. Danks, Gregory M. Shreve, Stephen B. Fountain and Michael K. McBeath. Thousand Oaks: Sage Publications, pp. 25–56.
- De Groot, Annette M. B. 2011. *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York and Hove: Psychology Press.
- Defrancq, Bart. 2015. Corpus-based research into the presumed effects of short evs. *Interpreting* 17: 26–45. [\[CrossRef\]](#)
- Defrancq, Bart, Koen Plevoets, and Cédric Magnifico. 2015. Connective items in interpreting and translation: Where do they come from? In *Yearbook of Corpus Linguistics and Pragmatics. Current Approaches to Discourse and Translation Studies*. Edited by Jesús Romero-Trillo. Heidelberg: Springer, vol. 3, pp. 195–222.
- Defrancq, Bart, and Koen Plevoets. 2018. Over-uh-load, filled pauses in compounds as a signal of cognitive load. In *Making Way in Corpus-Based Interpreting Studies*. Edited by Mariachiara Russo, Claudio Bendazzoli and Bart Defrancq. Singapore: Springer, vol. 1, pp. 43–64.
- Ferraresi, Adriano, and Silvia Bernardini. 2019. Building eptic: A many-sided, multi-purpose corpus of eu parliament proceedings. In *Parallel Corpora for Contrastive and Translation Studies: New Resources and Applications*. Edited by Irene Doval and Maria Teresa Sánchez Nieto. Amsterdam and Philadelphia: Benjamins, pp. 123–39.
- Gile, Daniel. 2009. *Basic Concepts and Models for Interpreter and Translator Training*, 2nd ed. Amsterdam: Benjamins.
- Grosjean, Francois. 2001. The bilingual's language modes. In *One Mind, Two Languages: Bilingual Language Processing*. Edited by Janet Nichol. Oxford: Blackwell, pp. 1–22.
- Gumul, Eva. 2006. Explicitation in simultaneous interpreting: A strategy or a by-product of language mediation? *Across Languages and Cultures. A Multidisciplinary Journal for Translation and Interpreting Studies* 7: 171–90. [\[CrossRef\]](#)
- Halverson, Sandra. 2014. Reorienting translation studies: Cognitive approaches and the centrality of the translator. In *Translation: A Multidisciplinary Approach*. Edited by Juliane House. New York: Palgrave Macmillan, pp. 116–39.
- Hansen, Sandra, and Silvia Hansen-Schirra. 2012. Grammatical shifts in english-german noun phrases. In *Cross-Linguistic Corpora for the Study of Translations*. Edited by Silvia Hansen-Schirra, Stella Neumann and Erich Steiner. Berlin: De Gruyter Mouton, pp. 133–45.
- Kajzer-Wietrzny, Marta, Adriano Ferraresi, Ilmari Ivaska, and Silvia Bernardini, eds. 2022a. *Empirical Investigations into the Forms of Mediated Discourse at the European Parliament*. Translation and Multilingual Natural Language Processing. Berlin: Language Science Press.
- Kajzer-Wietrzny, Marta, Adriano Ferraresi, Ilmari Ivaska, and Silvia Bernardini. 2022b. Using european parliament data in translation and interpreting research: An introduction. In *Empirical Investigations into the Forms of Mediated Discourse at the European Parliament*. Edited by Marta Kajzer-Wietrzny, Adriano Ferraresi, Ilmari Ivaska and Silvia Bernardini. Berlin: Language Science Press, pp. iii–xi.
- Koehn, Philipp. 2005. Europarl: A parallel corpus for statistical machine translation. Paper presented at MT Summit X, Phuket, Thailand, September 12–16. pp. 79–86.
- Lapshinova-Koltunski, Ekaterina, Christina Pollkläsener, and Heike Przybyl. 2022. Exploring explicitation and implicitation in parallel interpreting and translation corpora. *The Prague Bulletin of Mathematical Linguistics* 5: 5–22.

- Lapshinova-Koltunski, Ekaterina, Heike Przybyl, and Yuri Bizzoni. 2021. Tracing variation in discourse connectives in translation and interpreting through neural semantic spaces. Paper presented at the 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, November 7–11. Edited by Marine-Francine Moens, Xuanjing Huang, Lucia Specia and Scott Wen-tau Yih. Stroudsburg: Association for Computational Linguistics, pp. 134–42.
- Laviosa-Braithwaite, Sara. 2001. Universals of translation. In *Routledge Encyclopaedia of Translation Studies*. Edited by Mona Baker. London and New York: Routledge, pp. 288–91.
- Liu, Minhua. 2001. Expertise in Simultaneous Interpreting. A Working Memory Analysis. Ph.D. thesis, University of Texas, Austin, TX, USA.
- Liu, Minua, Ingrid Kurz, Barbara Moser-Mercer, and Miriam Shlesinger. 2020. The interpreter's aging: A unique story of multilingual cognitive decline? *Translation, Cognition & Behavior* 3: 287–310.
- Liu, Minhua, Diane L. Schallert, and Patrick J. Carroll. 2004. Working memory and expertise in simultaneous interpreting. *Interpreting* 6: 19–42. [\[CrossRef\]](#)
- Mauranen, Anna, and Pekka Kujamäki, eds. 2004. *Translation Universals: Do They Exist?* Amsterdam: Benjamins.
- Pastor, Gloria Corpas, and Fernando Sánchez Rodas. 2022. NLP-enhanced shift analysis of named entities in an English<>Spanish intermodal corpus of european petitions. In *Empirical Investigations into the Forms of Mediated Discourse at the European Parliament*. Edited by Marta Kajzer-Wietrzny, Adriano Ferraresi, Ilmari Ivaska and Silvia Bernardini. Berlin: Language Science Press, pp. 209–40.
- Przybyl, Heike, Alina Karakanta, Katrin Menzel, and Elke Teich. 2022. Exploring linguistic variation in mediated discourse: translation vs. interpreting. In *Empirical Investigations into the Forms of Mediated Discourse at the European Parliament*. Translation and Multilingual Natural Language Processing. Edited by Marta Kajzer-Wietrzny, Adriano Ferraresi, Ilmari Ivaska and Silvia Bernardini. Berlin: Language Science Press.
- Pym, Anthony. 2007. On shlesinger's proposed equalizing universal for interpreting. In *Interpreting Studies and Beyond: A Tribute to Miriam Shlesinger*. Edited by Franz Pöchhacker, Arnt Lykke and Inger M. Mees. Copenhagen: Samfundslitteratur Press, pp. 175–90.
- Serbina, Tatiana, Sven Hintzen, Paula Niemietz, and Stella Neumann. 2017. Changes of word class during translation—Insights from a combined analysis of corpus, keystroke logging and eye-tracking data. In *Empirical Modelling of Translation and Interpreting*. Edited by Silvia Hansen-Schirra, Oliver Čulo and Sascha Hofmann. Berlin: Language Science Press, pp. 177–208.
- Shlesinger, Miriam. 2008. Towards a definition of *Interpretes*: An intermodel, corpus-based study. In *Efforts and Models in Interpreting and Translation Research. A Tribute to Daniel Gile*. Edited by Gyde Hansen, Andrew Chesterman and Heidrun Gerzymisch-Arbogast. Amsterdam: John Benjamins, pp. 237–53.
- Shlesinger, Miriam, and Noam Ordan. 2012. More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target. International Journal of Translation Studies* 24: 43–60. [\[CrossRef\]](#)
- Steiner, Erich. 2001. Translations english–german: Investigating the relative importance of systemic contrasts and of the text-type “translation”. *SPRIKreports: Reports from the project Languages in Contrast* 7: 1–49.
- Steiner, Erich. 2012. A characterization of the resource based on shallow statistics. In *Cross-Linguistic Corpora for the Study of Translations*. Edited by Silvia Hansen-Schirra, Stella Neumann and Erich Steiner. Berlin: De Gruyter Mouton, pp. 71–89.
- Szmrecsanyi, Benedikt, and Alexandra Engel. 2022. Register variation in a cognitive (socio)linguistics perspective. In *Cognitive Sociolinguistics Revisited*. Edited by Gitte Kristiansen, Karlien Franco, Stefano De Pascale, Laura Rosseel, and Weiwei Zhang. Berlin and Boston: De Gruyter Mouton, pp. 398–409. [\[CrossRef\]](#)
- Teich, Elke. 2003. *Cross-Linguistic Variation in System and Text. A Methodology for the Investigation of Translations and Comparable Texts*. Berlin: Mouton de Gruyter.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. Paper presented at Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy, May 24–26. Luxembourg: European Language Resources Association (ELRA).
- Zellermayer, Michal. 1990. Shifting along the oral/literate continuum. *Poetics* 19: 341–357. [\[CrossRef\]](#)
- Čulo, Oliver, Silvia Hansen-Schirra, Stella Neumann, and Mihaela Vela. 2008. Empirical studies on language contrast using the english–german comparable and parallel CroCo Corpus. Paper presented at LREC 2008 workshop on “Building and Using COMPARABLE corpora”, Paris, France, May 28–30, Luxembourg: European Language Resources Association (ELRA), pp. 47–51.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.