PBML

The Prague Bulletin of Mathematical Linguistics
NUMBER 119 OCTOBER 2022 5–22

Exploring Explicitation and Implicitation in Parallel Interpreting and Translation Corpora

Ekaterina Lapshinova-Koltunski, Christina Pollkläsener, Heike Przybyl

Saarland University, Department of Language Science and Technology

Abstract

We present a study of discourse connectives in English-German and German-English translation and interpreting where we focus on the phenomena of explicitation and implicitation. Apart from distributional analysis of translation patterns in parallel data, we also look into surprisal, i.e. an information-theoretic measure of cognitive effort, which helps us to interpret the observed tendencies.

1. Introduction

The present paper deals with the phenomena of explicitation and implicitation (Klaudy and Károly, 2005; Blum-Kulka, 1986) in translation and interpreting. From the existing studies, we know that explicitation patterns differ in translation and interpreting (e.g. Lapshinova-Koltunski et al., 2021; Defrancq et al., 2015; Gumul, 2006, amongst others): while translation seems to share increased explicitness, interpreting rather shows a reduction of cohesive ties.

Existing corpus-based studies of explicitation and implicitation often looked into comparable data, contrasting distributions of discourse connectives in the subcorpora of translations and comparable non-translations in the target language (Puurtinen, 2004; Olohan and Baker, 2000). At the same time, detection of explicitation and implicitation effects is believed to require analysis of parallel corpora, i.e. looking at the aligned source texts and their translations (see e.g. Marco, 2018; Zufferey and Cartoni, 2014; Becher, 2011). In this study, we analyse parallel data to verify reported explicitation and implicitation trends in translation and interpreting. We inspect the

 Corresponding author: e.lapshinova@mx.uni-saarland.de
Cite as: Ekaterina Lapshinova-Koltunski, Christina Pollkläsener, Heike Przybyl. Exploring Explicitation and Implicitation in Parallel Interpreting and Translation Corpora. The Prague Bulletin of Mathematical Linguistics No. 119, 2022, pp. 5–22. doi: 10.14712/00326585.020.

translational pairs of discourse connectives in the sources and in the targets,[1] to detect explicitation patterns and strategies. We pay attention to the degree of the explicitation signal. Not all connectives have the same degree of how they signal a discourse relation (Crible, 2020; Asr and Demberg, 2012), as this depends on the number and frequency of other relations they may express. Number and frequency of discourse relations may also depend on the text register a connective occurs in (see details in Biber et al., 1999, p. 880ff). In general, ambiguous connectives express a weaker relation. For instance, a weak signal connective in the source, such as *aber* in example (1-a), being translated by a strong signal connective in the target, e.g. *however* in example (1-b), would indicate explicitation. No explicitation is observed if connectives hold a signal of the same degree, as *aber* transferred to *but* in interpreting in example (1-c).

(1) a. ***Aber** ich glaube, in einer Hinsicht gibt es Einigkeit...*(source)
b. ***However**, I believe that in one respect there is consensus...* (translation)
c. ***but** euh one thing we agree on...*(interpreting)

We analyse the distribution of the explicitation and implicitation cases for the same connectives in translation and interpreting in English and German to compare these phenomena across translation modes. We also seek for the explanation of these phenomena using the information-theoretical notion of surprisal, which indicates cognitive processing effort elicited in translation or interpreting.

We start from selected connectives, for which we reported explicitation and implicitation effects observed in bilingual semantic spaces in a previous study (Lapshinova-Koltunski et al., 2021). Relying on the connective lexicon Connective-Lex (Stede et al., 2019) and occurrences of connectives in a reference corpus (GECCo, Kunz et al., 2021), we estimate their signal strength. This is challenging, since we are looking at two languages and cross-lingual estimation of signal strength is not an easy task. We suggest a classification of connectives and their equivalents according to their explicitation or implicitation effects for both translation directions and explore the variation in these effects in translation and interpreting. We also pay attention to the type of relation the connectives express. Then, we look into the level of information conveyed by the connectives to interpret the results from a cognitive perspective.

The paper is structured as follows: in Section 2, we provide an overview of related studies, in Section 3, we present our methodology. The results are presented and discussed in Section 4. In Section 5, we conclude and give ideas for future work.

[1] In this way, our study is similar to translation spotting (Cartoni et al., 2013) a technique to disambiguate connective meaning, although we are not aiming at sense disambiguation but analyse transfer patterns in terms of explicitation or implicitation strategies.

2. Theoretical Background and Related Work

2.1. Explicitation and Implicitation

As already mentioned above, corpus-based analyses of explicitation and implicitation either look into comparable corpora, defining explicitation as a higher degree of cohesive explicitness in translations if compared to comparable non-translations (Puurtinen, 2004; Olohan and Baker, 2000), or analyse parallel data to uncover transformations from the source text to the target (Marco, 2018; Zufferey and Cartoni, 2014). Explicitation or implicitation effects are often related to the increased usage of discourse connectives and have been extensively analysed so in both human and machine translation (Shi et al., 2019; Meyer and Webber, 2013; Hoek and Zufferey, 2015; Hoek et al., 2015, amongst others). Discourse connectives have also been addressed within the studies on interpreting. For instance, Gumul (2006, p. 184) stated that explicitation in interpreting is related to adding discourse markers among other means of cohesive explicitness. At the same time, Shlesinger (1995) observed a reduction of cohesive ties in interpreting if compared to the source language input (implicitation). And Kajzer-Wietrzny (2012) showed that there are differences between translation and interpreting in the usage of linking adverbials, with translation being more explicit. Defrancq et al. (2015) found that interpreters reshaped the discourse structure of the source speeches in terms of connectives. For our research purposes, we adopt the definition of explicitation introduced by Klaudy and Károly (2005, p. 15): explicitation takes place when a translation contains more specific linguistic units instead of more general units in the source, or new linguistic units not present in the source. Previous studies (Przybyl et al., 2022a) also show that in marking logical relations, interpreters tend to prefer more general items over more specific ones (e.g. *but* vs. *however*), which is typical of spoken production (Crible and Cuenca, 2017), and use fewer different, but polyfunctional discourse markers.

Explicitation and implicitation effects also depend on the type of relations discourse connectives trigger: cognitively simple relations are more often left implicit than relations that are cognitively more complex (see Hoek et al., 2017). This is also confirmed in a recent study by Blumenthal-Dramé (2021) who showed that the processing of concessive sentences benefits more from the explicit marking than the processing of causal sentences, as causal links are more expected than concessive ones. Hoek et al. (2015) also show that explicitation and implicitation maybe affected by expectedness of discourse relations, as defined on the basis of the continuity hypothesis (Murray, 1997) and the causality-by-default hypothesis (Sanders, 2005). In our study, we also look at (un)expectedness of discourse relations via connectives as indexed by surprisal (see Section 2.2 below). However, in our previous study on translation and interpreting (Lapshinova-Koltunski et al., 2021), we could not find the effect of relation type on the explicitation and implicitation in the analysed interpreting data, as the analysed neural semantic spaces in interpreting contained more implicitation than translation independent of the relation the connectives triggered.

The real driving force of explicitation is not easy to determine as many factors at once may be involved. We hypothesize that explicitation, on the one hand, facilitates processing for the producers (translator/interpreter), and on the other hand, helps to shape the content for the recipient (audience design). We aim to inspect explicitation through discourse connectives in both comparable corpora – translation/interpreting and comparable originals of the target language, and parallel corpora – the aligned source language inputs and target language outputs. We expect the parallel data to show (1) if discourse connectives are used in translation/interpreting simply because the source texts already contain such items and they are transferred into the target (equivalence); (2) if translators/interpreters leave them out or change them from more specific to more general, e.g. *however* to *but* (implicitation); if translators/interpreters change more general items to more specific ones (explicitation). We assume that from the cognitive perspective, equivalence and implicitation occur to facilitate processing for translators or interpreters. At the same time, implicitation in interpreting is used due to time pressure, which is usually not the case in translation. Explicitation is used to better shape the content for the audience.

2.2. Surprisal

Apart from the distributional analysis of discourse connectives, we also use a probabilistic measure based on Information Theory (Shannon, 1948), i.e. surprisal or *unpredictability in context*. Surprisal adds a direct link to cognition (Teich et al., 2020) as it represents a direct indicator for cognitive processing effort elicited, in our cases by either the source text or the translation output. Surprisal is a word-based measure of cognitive effort (Hale, 2001, p. 4), i.e. highly predictable words, that incur low surprisal, require low cognitive processing effort. Surprisal measures help to analyse language use in terms of rational communication, and account for a trade-off between expressiveness and efficiency.

In translation, and especially in interpreting, cognitive constraints (e.g. time pressure) impact cognitive processing, and as a consequence, language shape. Surprisal measures are assumed to shed a light on such constraints. Therefore, we use surprisal to investigate the driving force of explicitation and implicitation attempting to link them to cognitive processing. A few studies have already used similar information-theoretic measures (entropy, perplexity) to compare translated and non-translated texts (Bizzoni and Lapshinova-Koltunski, 2021; Teich et al., 2020; Martínez and Teich, 2017; Rubino et al., 2016). Mostly, these measures were used for a comparable analysis of texts, i.e. translated or interpreted texts and comparable originals of the target language. Some studies (Martínez and Teich, 2017; Schaeffer et al., 2015) also used word translation entropy indicating how many equally likely translations may be produced for a source word in a given context. Higher translation entropy means more lexical choices for the translator and higher cognitive effort on the translator's side. This was confirmed by Schaeffer et al. (2015) with experimental data.

In our study, we look at surprisal of discourse connectives in the source texts, as well as in translated and interpreted texts. We also compare surprisal of the same connectives across translation and interpreting. We assume that connectives with low surprisal in the source texts would require low cognitive effort for a translator or an interpreter as a recipient. Target side surprisal indicates cognitive effort of a recipient of the translated or interpreted texts (reader or listener). As ambiguous connectives offer more translation options, they require a higher cognitive effort for a translator as a transmitter. We expect that translators and interpreters use equivalence or implicitation strategies defined in Section 2.1 above more often in case of ambiguous connectives, as explicitation strategies would require more cognitive effort of the translator/interpreter. We also assume that surprisal of the same connectives would vary in translation and interpreting, as the interpreting environment poses more constraints.

3. Methodology

3.1. Corpus

As data we use the bidirectional, sentence-aligned corpus versions of Europarl-UdS (Karakanta et al., 2018) and EPIC-UdS (Przybyl et al., 2022b),[2] specifically the English↔German supcorpora. Europarl-UdS includes the officially published original speeches held at the European Parliament, as well as their translations. EPIC-UdS is the spoken counterpart, consisting of transcripts of these speeches and their simultaneous interpretation, without any corrections with respect to the spoken signal. The total number of tokens comprises approx. 26 millions (more details on subcorpora size are given in Table 1 in Appendix). Both corpora are sentence-aligned and include rich annotation (lemma, part-of-speech, dependencies, surprisal). The absolute number of the extracted and analysed instances of connectives amounts to 87366 and 1242 for the written and spoken corpora respectively.

3.2. Methods

We classified different translation options for each connective into three groups, depending on which translation strategy (explicitation, implicitation and equivalence) was used (see Section 2.1 above). This was not a trivial task as meaning, function, distribution and stylistic preferences almost never display one-to-one correspondence in two languages. Relying on the assumption that the degree of how connectives signal a discourse relation depends on number, frequency and text register preferences of the relations, we combined several approaches to decide if translation options were equivalent, explicit or implicit compared to the source. First, we used a German-English bilingual dictionary (Deuter, 2019) to look up the semantically equivalent

[2] http://hdl.handle.net/21.11119/0000-0008-F519-8

connectors for each language. Second, we checked if the logical relations signalled by the German and English connectives overlapped, using the web-based multilingual lexical resource Connective-Lex.[3] Thirdly, we used the GECCo corpus (Kunz et al., 2021), which is annotated with conjunctive relations and contains different spoken and written registers, to compare the distribution among registers for the different connectives. For example, Deuter (2019) lists *falls* as equivalent to *if*. However, the other resources show that *falls* signals fewer logical relations and appears in fewer registers compared to *if*. That means that *falls* can be used in considerably fewer contexts than *if*, which is why is was classified as more explicit. Although a certain amount of subjectivity cannot be denied in this approach, we consider it reliable, as all the three linguists involved in this study agreed on the underlying classification.

We use the tool CQPweb (Hardie, 2012) to extract instances of connectives along with their distributional information from parallel data. The CQP query language allows to include restrictions on the searched structures. We queried for source connectives and looked at the alignments to identify the most frequent corresponding connectives in the target. Then, we queried for these source-target-combinations and extracted frequencies. Once all frequent translation options were identified, we extracted all instances of the source connective without corresponding target connective (*implicit_none*). This does not mean that there is no signal at all in the target. Instead, it means that we consider overt connectives to be stronger signals of logical relations compared to other types of signals like the V1-syntactic construction in example (2).

(2) ***If** we had more cooperation with the Member States, we would not have as many problems* (source). - ***Würde** die Zusammenarbeit mit den Mitgliedstaaten besser funktionieren, hätten wir nicht so viele Probleme* (translation).

Some noise in the data was inevitable. We tried to fine-grain some queries as to exclude non-connective usages. For example, we removed *because of* and only queried for sentence-initial *also*. However, to remove all noise, a manual investigation of the data would have been necessary, which would have been outside the scope of this paper and its exploratory aim.

As mentioned in Section 2.2, to estimate the level of information conveyed by the connectives, we use the information theoretical notion of surprisal. Surprisal measures information content of a word (w) in number of bits, calculated as the negative log base 2 probability of the word in context. Context is defined here as three preceding words, cf. Equation (1).

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

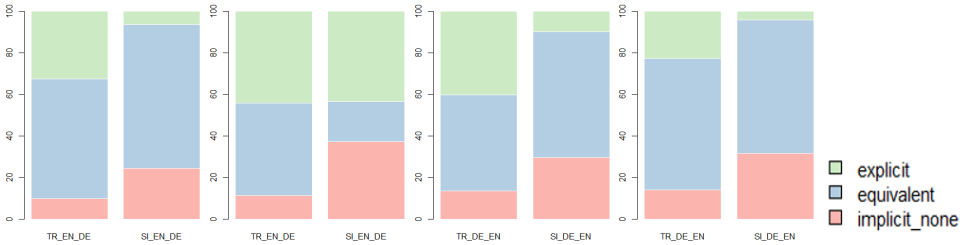
Every word in our corpora is annotated with its surprisal value. We extract the surprisal values for every connective analysed.

[3] http://connective-lex.info/

4. Results

4.1. Translation equivalents

We analyse the frequency distributions of translations as described in Section 3.2 above, paying attention to the translation patterns. An overview of all analysed connectives is given in Table 2 in the Appendix. Due to space restrictions, we visualise translations (explicit=explicitation, equivalent=equivalence, implicit_none=implicitation, i.e. leaving out a connective) of four connectives with stacked bar plots (Figure 1). Generally, we see that there is more explicitation in translation (TR) and more implicitation without overt connectives in interpreting (SI). Equivalence is equally used in both translation modes, with an exception of translation of *because*. In interpreting, we observe more implicitation than equivalence as well as an unusual amount of explicitation for this connective.[4] The amount of explicitation of *because* is the same in both translation modes.



(a) but (b) because (c) aber (d) weil

Figure 1. Distribution for translation/interpreting of but, because, aber and weil

4.2. Surprisal

Now we look into the cognitive processing effort elicited by connectives. First, we compare surprisal of the same connectives in original written and spoken production to translations and interpreting to see if their predictability in context, and thus the cognitive processing load involved, differs within the written and spoken mode. Figures 2 and 3 visualise surprisal values for the four selected connectives in comparable originals (ORG) and translation/interpreting (TR/SI) of the same language. As corpus size and transcription differences (e.g. punctuation in written, no punctuation

[4]The reported differences for the connectives used in translation vs interpreting are confirmed by a Pearson's Chi-squared test: *but*, *because* and *aber* p-value < 2.2e-16, *weil* p-value = 3.965e-08.

in spoken) have an influence on surprisal, we cannot compare written and spoken surprisal values directly.

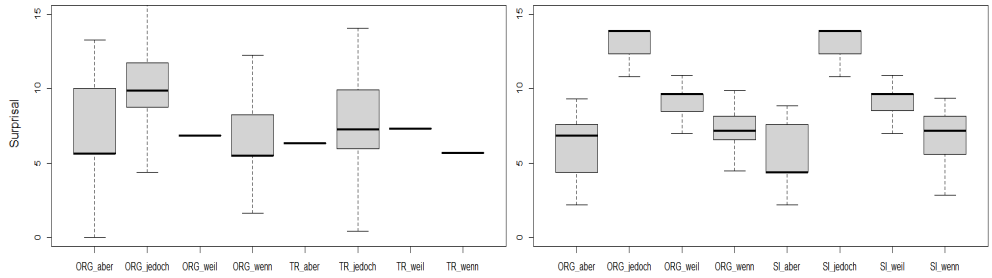


Figure 2. Surprisal in German texts in written (left) and spoken (right) mode

Overall we can observe that for the written mode, most connectives are predictable to the same degree in the given context (visualised by all boxplot quartiles being very close to the median and therefore a line is displayed instead of a box, or narrow boxes with no or short whiskers), which means that processing effort linked to these connectives does not depend on the text production type (translated or not). Translations show less variation concerning surprisal of connectives than written originals, with the exception of *jedoch*. Interestingly, this is the only case where surprisal of a connective is lower in translation than in comparable originals (its processing requires lower cognitive effort). In spoken data, a similar case is observed for *aber* – its surprisal and hence, cognitive processing, is lower in interpreting than in original speech.

In general, surprisal varies more in the spoken data,[5] both in German and English. We do not observe the same tendencies for *however* (equivalent of *jedoch*) in the written and *but* in the spoken data in English). None of the connectives have lower surprisal in either translation or interpreting. Instead, we observe higher surprisal for *because* and *however* in interpreting, which means that they are less expected in interpreting and their occurrence requires a higher processing effort. Comparing surprisal of individual connectives, we observe statistical differences for all connectives studied in both languages for the written mode.[6] Differences are marginal for the spoken mode, with no statistical effect.

We also analyse the parallel data with the focus on the information content of equivalent and explicit use of connectives in the target to see if the strategies used provide any processing bonus for the audience. As stated in Section 2.2, while equiv-

[5]However, differences between the spoken and written mode might be due to mid sentence punctuation being included as context in the written mode which are not transcribed for the spoken data.

[6]Wilcoxon rank sum test with continuity correction for non parametric data: p-value < 2.2e-16.

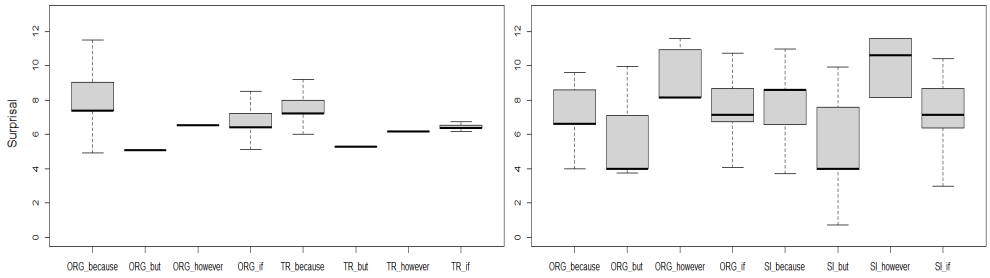
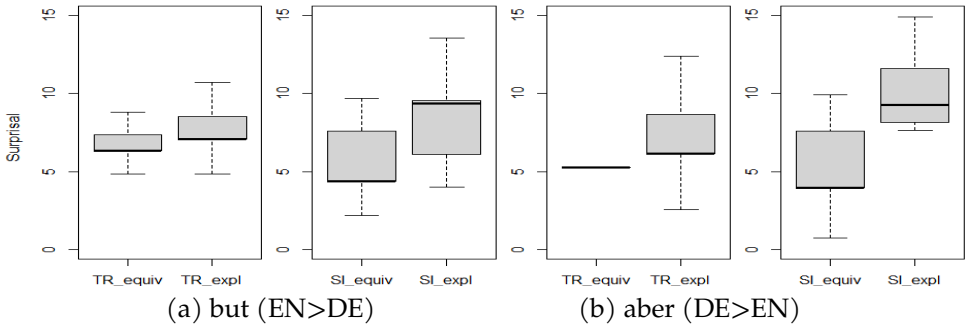


Figure 3. Surprisal in English texts in written (left) and spoken (right) mode

alence is used to reduce the effort on the translator's side, explicitation would mean higher cognitive effort for the translator. However, explicitation in this case could be used for the sake of audience design (to reduce the effort of the reader/listener). Surprisal values are used to compare cognitive processing effort caused by translation patterns (equivalent or explicitation).[7] We extract surprisal of the translations of specific connectives and summarise the surprisal values according to the strategies used. We are not able to analyse processing effort in the cases of implicitation where a connective is left out in the target, as surprisal is calculated for words and we cannot calculate surprisal of a zero.



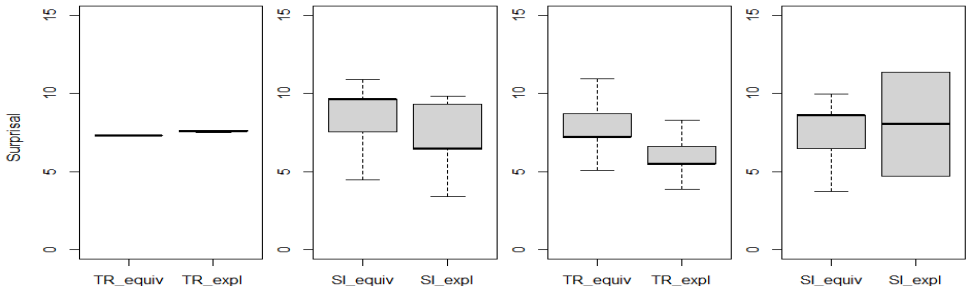
(a) but (EN>DE) (b) aber (DE>EN)

Figure 4. Surprisal for translations of but and aber in in translation and interpreting

We observe differences in processing explicitation across languages for translation of contrast relations triggered by *but* and *aber* in the source (Figure 4): in both trans-

[7]Note that we do not compare surprisal cross-lingually, i.e. between the source connectives and their translations.

lation directions and both modes, explicitation causes a higher cognitive processing effort. The distance between surprisal of equivalent and explicit targets is greater in interpreting than in translation and therefore explicitation causes an even higher processing load for the recipient of the spoken message than the written one. The observed differences are significant for both languages and modes[8].



(a) because (EN>DE) (b) weil (DE>EN)

Figure 5. Surprisal for translations of because and weil in translation and interpreting

For translation of contingency relations triggered by *because* and *weil* in the source, we observe a different tendency (Figures 5). In general, explicitation seems to provide a processing bonus for both translation and interpreting: it either causes a similar processing load as equivalence (German translation and English interpreting)[9] or it even causes a lower processing effort than equivalence (in German interpreting and English translation). This means that processing of this contingency discourse connective benefits more from the explicit marking in both translation modes, and even more in German interpreting and English translation in our data.

This result could appear controversial to the studies showing that cognitively simple relations (e.g. the relation of contingency) are more expected and thus, are more often left implicit than more complex ones (such as comparison). However, in this analysis we miss the cases of explicitation from zero connectives in the source, as well as cases of implicitation to zeros in the target, which does not allow us to report a more comprehensive account of the translation patterns for different relation types.

[8]Wilcoxon rank sum test with continuity correction for non parametric data, TR *but* and *aber*: p-value < 2.2e-16; SI *but*: p-value=4.036e-07, SI *aber*: p-value=2.495e-13.

[9]The reported differences are significant for all subsets apart from for *weil* in the spoken mode. Wilcoxon rank sum test with continuity correction for non parametric data, TR *because* and *weil*: p-value < 2.2e-16; SI *because*: p-value = 0.0004957; SI *weil*: not significant.

5. Conclusion, Discussion and Future Work

The present paper deals with translation of connectives involving explicitation and implicitation effects in spoken and written data. We describe various patterns or strategies for a selected number of connectives for English-German and German-English translation and interpreting, reporting on the distributional preferences across the language pairs. We show that equivalence and implicitation are more frequently used in interpreting than translation, as these strategies facilitate cognitive processing in high-time-pressure situations. Moreover, we analysed surprisal of connectives in comparable source and target context. We showed that the same connectives convey a similar level of information, and hence cognitive load, in written originals and translations in the same language for both modes, with some connectives being less expected in interpreting. We also analysed surprisal of translation patterns to discover that explicitation, although challenging for translator or interpreter in case of ambiguous connectives, may provide a bonus in cognitive processing effort for the recipient, especially in case of translation and interpreting of contingency relations. In this case, as assumed, explicitation is used to shape the content for the audience.

In future, we would like to look into cognitive load conveyed by explicitation, when the source does not contain any signal of a relation and a connective is added in translation/interpreting. As our current approach to surprisal calculation has some limitations (calculated on the word level, depending on word sequences), we plan to use a different surprisal measure, i.e. average surprisal of sentences. First of all, this will be more appropriate for the analysis of connectives, as connective usage depends on relations between clauses and sentences, which is out of the scope of the current surprisal calculation. Moreover, comparing average surprisal of sentences with or without a connective would also compensate for the implicitation cases not covered by the current analysis – with the approach used, we are not able to analyse the cases with an omitted connective in translation/interpreting, as surprisal values were calculated on the level of words. Besides, we will explore interaction of such cross-linguistic constraints as function and stylistic preferences and their impact on the transformation patterns.

Acknowledgements

This paper is based on research conducted in a project funded by the Deutsche Forschungsgemeinschaft – SFB 1102/Project B7, ID 232722074.

Bibliography

Asr, Fatemeh Torabi and Vera Demberg. Measuring the Strength of Linguistic Cues for Discourse Relations. In *Proceedings of the Workshop on Advances in Discourse Analysis and its Computational Aspects*, pages 33–42, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL https://aclanthology.org/W12-4703.

Becher, Viktor. *Explicitation and implicitation in translation. A corpus-based study of English-German and German-English translations of business texts*. PhD thesis, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky, 2011.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. *Longman Grammar of Spoken and Written English*. Longman, 1999.

Bizzoni, Yuri and Ekaterina Lapshinova-Koltunski. Measuring Translationese across Levels of Expertise: Are Professionals more Surprising than Students? In *Proceedings of the 23rd NoDaLiDa*, pages 53–63, Online, May 31 - June 02 2021. ACL. URL https://aclanthology.org/2021.nodalida-main.6.

Blum-Kulka, Shoshana. Shifts of Cohesion and Coherence in Translation. In House, Juliane and Shoshana Blum-Kulka, editors, *Interlingual and intercultural communication*, pages 17–35. Gunter Narr, Tübingen, 1986.

Blumenthal-Dramé, Alice. The Online Processing of Causal and Concessive Relations: Comparing Native Speakers of English and German. *Discourse Processes*, 0(0):1–20, 2021. doi: 10.1080/0163853X.2020.1855693. URL https://doi.org/10.1080/0163853X.2020.1855693.

Cartoni, Bruno, Sandrine Zufferey, and Thomas Meyer. Annotating the meaning of discourse connectives by looking at their translation: The translation-spotting technique. *Dialogue and Discourse*, 4:65–86, 2013. doi: 10.5087/dad.2013.204.

Crible, Ludivine. Weak and strong discourse markers in speech, chat and writing: Do signals compensate for ambiguity in explicit relations? *Discourse Processes*, 2020. doi: 10.1080/0163853X.2020.1786778.

Crible, Ludivine and Maria Josep Cuenca. Discourse markers in speech: Characteristics and challenges for corpus annotation. *Dialogue and Discourse*, 8:149–166, 2017.

Defrancq, Bart, Koen Plevoets, and Cédric Magnifico. Connective Items in Interpreting and Translation: Where Do They Come From? *Yearbook of Corpus Linguistics and Pragmatics*, 3:195–222, 2015. doi: 10.1007/978-3-319-17948-3_9.

Deuter, Margaret, editor. *Das Große Oxford Wörterbuch: Englisch-Deutsch / Deutsch-Englisch*. Oxford University Press, third edition, 2019.

Gumul, Ewa. 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*, 7:171–190, 2006. doi: 10.1556/Acr.7.2006.2.2.

Hale, John. 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, 2001. Association for Computational Linguistics. doi: 10.3115/1073336.1073357.

Hardie, Andrew. CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17:380–409, 2012. doi: 10.1075/ijcl.17.3.04har.

Hoek, Jet and Sandrine Zufferey. Factors Influencing the Implicitation of Discourse Relations across Languages. In *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*, London, UK, April 2015. Association for Computational Linguistics. URL https://aclanthology.org/W15-0205.

Hoek, Jet, Jacqueline Evers-Vermeul, and Ted J.M. Sanders. The Role of Expectedness in the Implicitation and Explicitation of Discourse Relations. In *Proceedings of the Second Workshop on Discourse in Machine Translation*, pages 41–46, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/W15-2505. URL https://aclanthology.org/W15-2505.

Hoek, Jet, Sandrine Zufferey, Jacqueline Evers-Vermeul, and Ted J.M. Sanders. Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics*, 121:113–131, 2017. doi: 10.1016/j.pragma.2017.10.010.

Kajzer-Wietrzny, Marta. *Interpreting universals and interpreting style*. PhD thesis, Uniwersytet im. Adama Mickiewicza, Poznan, Poland, 2012. Unpublished PhD thesis.

Karakanta, Alina, Mihaela Vela, and Elke Teich. Europarl-UdS: Preserving Metadata from Parliamentary Debates. In Fišer, Darja, Maria Eskevich, and Franciska de Jong, editors, *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), 2018.

Klaudy, Kinga and Krisztina Károly. Implicitation in Translation: Empirical Evidence for Operational Asymmetry in Translation. *Across Languages and Cultures*, 6:13–28, 2005. doi: 10.1556/Acr.6.2005.1.2.

Kunz, Kerstin, Ekaterina Lapshinova-Koltunski, José Manuel Martínez Martínez, Katrin Menzel, and Erich Steiner. *GECCo – German-English Contrasts in Cohesion: Insights from Corpus-based Studies of Languages, Registers and Modes*, volume 355 of *Trends in Linguistics. Studies and Monographs* [*TiLSM*]. De Gruyter Mouton, 2021. ISBN 9783110711073. doi: 10.1515/9783110711073.

Lapshinova-Koltunski, Ekaterina, Heike Przybyl, and Yuri Bizzoni. Tracing variation in discourse connectives in translation and interpreting through neural semantic spaces. In *Proceedings of CODI at EMNLP-2021*, pages 134–142, Punta Cana and Online, 10–11 November 2021. ACL. doi: 10.18653/v1/2021.codi-main.13.

Marco, Josep. Connectives as indicators of explicitation in literary translationA study based on a comparable and parallel corpus. *Target*, 30:87–111, 2018. doi: 10.1075/target.16042.mar.

Martínez, José Manuel Martínez and Elke Teich. Modeling Routine in Translation with Entropy and Surprisal: A Comparison of Learner and Professional Translations. In Cercel, Larissa, Marco Agnetta, and Maria Teresa Amido Lozano, editors, *Kreativität und Hermeneutik in der Translation*. Narr Francke Attempto Verlag, 2017.

Meyer, Thomas and Bonnie Webber. Implicitation of Discourse Connectives in (Machine) Translation. In *Proceedings of the Workshop on Discourse in Machine Translation*, pages 19–26, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL https://aclanthology.org/W13-3303.

Murray, John D. Connectives and narrative text: The role of continuity. *Memory and Cognition*, 25:227–236, 1997. doi: 10.3758/BF03201114.

Olohan, Maeve and Mona Baker. Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures*, 1:141–158, 2000. doi: 10.1556/Acr.1.2000.2.1.

Przybyl, Heike, Alina Karakanta, Katrin Menzel, and Elke Teich. Exploring linguistic variation in mediated discourse: translation vs. interpreting. In Kajzer-Wietrzny, Marta, Silvia Bernardini, Adriano Ferraresi, and Ilmari Ivaska, editors, *Empirical investigations into the forms of mediated discourse at the European Parliament*, Translation and Multilingual Natural Language Processing. Language Science Press, Berlin, 2022a.

Przybyl, Heike, Ekaterina Lapshinova-Koltunski, Katrin Menzel, Stefan Fischer, and Elke Teich. EPIC UdS - Creation and Applications of a Simultaneous Interpreting Corpus. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pages 1193–1200, Marseille, France, 20-25 June 2022b. ELDA. URL https://aclanthology.org/2022.lrec-1.127.

Puurtinen, Tiina. Explicitation of clausal relations: A corpus-based analysis of clause. Connectives in translated and non-translated Finnish children's literature. In Mauranen, A. and P. Kujamäki, editors, *Translation universals: Do they exist?*, pages 165–76. John Benjamins, Amsterdam/Philadelphia, 2004. doi: 10.1075/btl.48.13puu.

Rubino, Raphael, Ekaterina Lapshinova-Koltunski, and Josef van Genabith. Information Density and Quality Estimation Features as Translationese Indicators for Human Translation Classification. In *Proceedings of NAACL HT 2006*, pages 960–970, San Diego, California, June 12-17 2016. doi: 10.18653/v1/N16-1110.

Sanders, Ted J. M. Coherence, causality and cognitive complexity in discourse. In *Proceedings/Actes SEM-05, First International Symposium on the Exploration and Modelling of Meaning*, pages 105–114, 2005.

Schaeffer, Moritz, Barbara Dragsted, Kristian Tangsgaard Hvelplund, Laura Winther Balling, and Michael Carl. Word Translation Entropy: Evidence of Early Target Language Activation During Reading for Translation. In *Technical Committee on Thought and Language (TL)*, Tokyo, Japan, 2015. doi: 10.1007/978-3-319-20358-4_9. URL http://www.ieice.org.

Shannon, Claude E. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

Shi, Weijia, Muhao Chen, Yingtao Tian, and Kai-Wei Chang. Learning Bilingual Word Embeddings Using Lexical Definitions. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*, pages 142–147, Florence, Italy, August 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-4316. URL https://aclanthology.org/W19-4316.

Shlesinger, Miriam. Shifts in cohesion in simultaneous interpreting. *The Translator*, 1:193–214, 1995. doi: 10.1080/13556509.1995.10798957.

Stede, Manfred, Tatjana Scheffler, and Amália Mendes. Connective-Lex: A Web-Based Multilingual Lexical Resource for Connectives. *Discours*, 24, 2019. doi: 10.4000/discours.10098.

Teich, Elke, José Martînez Martînez, and Alina Karakanta. Translation, information theory and cognition. In Alves, Fabio and Arnt Lykke Jakobsen, editors, *The Routledge Handbook of Translation and Cognition*, chapter 20. Routledge, London, 2020. ISBN 9781138037007. doi: 10.4324/9781315178127-24.

Zufferey, Sandrine and Bruno Cartoni. A multifactorial analysis of explicitation in translation. *Target*, 26(3):361–384, 2014. doi: 10.1075/target.26.3.02zuf.

Appendix

Europarl UdS	tokens	EPIC UdS	tokens
ORG WR EN	8,693,135	ORG SP EN	67,526
TR EN DE	3,100,647	SI EN DE	57,532
ORG WR DE	7,869,289	ORG SP DE	56,488
TR DE EN	6,260,869	SI DE EN	58,503

Table 1. Corpus overview EUROPARL UdS (written) and EPIC UdS (spoken)

Source	Target	Category	fpm TR	fpm SI
		CONTINGENCY		
if	wenn	equivalent	1,482.27	1,585.48
if	falls	explicit	75.51	30.20
if	ob	equivalent	113.69	45.30
if	none	implicit_none	490.70	875.79
wenn	if	equivalent	787.00	1,831.08
wenn	when	equivalent	32.40	35.55
wenn nicht	unless	equivalent	13.98	0.00
wenn nicht	albeit	explicit	678.23	408.88
auch wenn	even though	equivalent	56.28	17.78
wenn	none	implicit_none	705.69	1,084.43
because	weil	equivalent	574.69	423.00
because	denn	explicit	407.83	921.00
because	da	explicit	161.49	30.00
because	none	implicit_none	147.00	815.39
weil	because	equivalent	852.69	1,084.43
weil	since	explicit	53.89	35.55
weil	for	explicit	77.25	0.00
weil	as	explicit	175.83	35.55
weil	none	implicit_none	191.52	533.33
		TEMPORAL		
finally	schließlich	equivalent	79.76	60.40
finally	abschließend	explicit	104.08	0.00
finally	zu dem Schluss	explicit	14.99	15.10

continued on next page

continued from previous page

Source	Target	Category	fpm TR	fpm SI
finally	zu dem Abschluss	explicit	7.35	15.10
finally	zu guter letzt	explicit	0.85	15.10
finally	dann	implicit	7.92	60.40
finally	letztlich / letztendlich	explicit	7.35	15.10
finally	ein/als letztes	explicit	0.57	0.00
finally	letztens	explicit	2.26	0.00
finally	letzter Punkt / letzte Anmerkung / letzte Bemerkung / letztes Wort	explicit	2.55	0.00
finally	nicht zuletzt	explicit	2.55	0.00
finally	none	implicit_none	92.77	135.90
schließlich	finally	equivalent	34.62	17.78
schließlich	ultimately	equivalent	10.06	0.00
schließlich	in the end	explicit	3.24	0.00
schließlich	after all	explicit	21.32	0.00
schließlich	at the end of the day	explicit	6.48	0.00
schließlich	lastly	explicit	3.58	0.00
schließlich	at last	explicit	0.85	0.00
schließlich	last but not least	explicit	0.85	0.00
schließlich	in the final analysis	explicit	2.05	0.00
schließlich	in conclusion	explicit	0.85	0.00
schließlich	none	implicit_none	20.29	17.78
	EXPANSION			
also	außerdem	equivalent	7.64	0.00
also	auch	equivalent	4.81	45.30
also	ebenfalls	explicit	1.98	0.00
also	ebenso	explicit	1.13	0.00
also	darüber hinaus	explicit	3.11	0.00
also	gleichfalls	explicit	0.28	0.00
also	ferner	explicit	3.11	0.00
also	des Weiteren	explicit	0.57	0.00
also	zudem	explicit	3.39	0.00
also	und	implicit	1.98	30.20
also	hinzu kommt	explicit	1.13	0.00
also	zusätzlich	explicit	0.28	0.00
also	none	implicit_none	6.50	15.10

continued on next page

continued from previous page

Source	Target	Category	fpm TR	fpm SI
außerdem	also	equivalent	47.92	35.55
außerdem	furthermore	equivalent	7.50	0.00
außerdem	moreover	equivalent	17.22	0.00
außerdem	in addition	explicit	19.95	0.00
außerdem	what is more	explicit	1.19	0.00
außerdem	besides	equivalent	1.88	0.00
außerdem	apart from that	explicit	1.02	0.00
außerdem	none	implicit_none	12.79	35.55
		COMPARISON		
but	aber	equivalent	1,558.07	3,503.16
but	sondern	equivalent	597.89	392.60
but	jedoch	explicit	393.97	0.00
but	doch	explicit	684.71	286.90
but	allerdings	explicit	83.43	60.40
but	dennoch	explicit	52.32	15.10
but	none	implicit_none	374.17	1,374.08
aber	but	equivalent	2,007.75	2,328.85
aber	however	explicit	1,081.39	248.88
aber	though	explicit	256.83	53.33
aber	although	explicit	139.67	0.00
aber	while	explicit	113.24	17.78
aber	whilst	explicit	35.64	0.00
aber	nevertheless	explicit	43.66	17.78
aber	yet	equivalent	59.18	0.00
aber	nonetheless	explicit	29.67	35.55
aber	none	implicit_none	598.08	1,137.76
however	aber	implicit	109.45	181.20
however	jedoch	equivalent	486.74	15.10
however	doch	implicit	89.09	0.00
however	allerdings	equivalent	156.40	30.20
however	dennoch	equivalent	36.20	0.00
however	none	implicit_none	128.97	45.30
jedoch	however	equivalent	66.17	35.55
jedoch	but	implicit	142.74	17.78
jedoch	though	implicit	21.32	0.00
jedoch	although	implicit	17.74	0.00
jedoch	while	explicit	11.77	0.00
jedoch	whilst	explicit	2.56	0.00

continued on next page

continued from previous page

Source	Target	Category	fpm TR	fpm SI
jedoch	nevertheless	explicit	5.63	0.00
jedoch	yet	implicit	7.84	0.00
jedoch	nonetheless	explicit	2.39	0.00
jedoch	none	implicit_none	36.50	17.78

Table 2. Distribution of connectives and their translations: fpm=frequency per million, TR=translation, SI=interpreting

Address for correspondence:
Ekaterina Lapshinova-Koltunski
e.lapshinova@mx.uni-saarland.de
Department of Language Science and Technology
Campus A2 2, Room 1.02
D-66123 Saarbrücken, Germany