Using Translation Process Data to Explore Explicitation and Implicitation through Discourse Connectives

Ekaterina Lapshinova-Koltunski
University of Hildesheim
Lübecker Str. 3
DE-31141 Hildesheim
lapshinovakoltun@uni-hildesheim.de

Michael Carl
Kent State University
800 E. Summit St.
Kent, OH 44242
mcarl6@kent.edu

Abstract

We look into English-German translation process data to analyse explicitation and implicitation phenomena of discourse connectives. For this, we use the database CRITT TPR-DB which contains translation process data with various features that elicit online translation behaviour. We explore the English-German part of the data for discourse connectives that are either omitted or inserted in the target, as well as cases when changing a weak signal to strong one, or the other way around. We determine several features that have an impact on cognitive effort during translation for explicitation and implicitation. Our results show that cognitive load caused by implicitation and explicitation may depend on the discourse connectives used, as well as on the strength and the type of the relations the connectives convey.

1 Introduction

Explicitation in translation is often defined as an increased usage of linking devices, such as discourse connectives. Implicitation is an opposite phenomenon and means a decrease in the number of connectives used in translation because of frequent ommissions. Both explicitation and implicitation belong to the phenomena of translationese (Gellerstam, 1986; Baker, 1993; Toury, 1995, amongst others). The latter have received an increased attention in multilingual language processing (see e.g. Dutta Chowdhury et al., 2020; Artetxe et al., 2020; Graham et al., 2020). In this paper, we analyse explicitation and implicitaion phenomena from a cognitive perspective, i.e. looking into translation process data. The data under analysis is parallel, so that we are able to inspect the translational pairs of English discourse connectives in the sources and their translations into German. Apart from taking into consideration omission or insertion of a connective, we also analylse transformation cases, when the degree of the explicitation signal is changed. The strength of the signal a connective conveys depends on the number and frequency of relations they may trigger (Asr and Demberg, 2012; Crible, 2020): ambiguous connectives convey a weaker signal. We interpret translation from a weak signal connective, e.g. *but* in example (1-a) into a strong signal connective, e.g. *jedoch* in example (1-b), as explicitation. No explicitation (equivalence) is observed if connectives hold a signal of the same degree: *but* translated into *aber* in example (1-c).

(1) a. *Some of the most vulnerable countries of the world have contributed the least to climate change, but are bearing the brunt of it.*
b. *Einige der Länder, die weltweit am wenigsten zum Klimawandel beigetragen, tragen jedoch die Hauptlast.*
c. *Einige der am meisten gefährdeten Länder der Welt haben am wenigsten zum Klimawandel beigetragen, leiden aber dessen Folgen.*

We start from the general cases of implicitation and explicitation (tokens marked by a syntactic parser as connectives left out or added in translation) and analyse three features of describing behaviour during translation: production (typing) pauses and reading time in translation unit. We also look at the probability of translation choices. Then, we analyse translation patterns of the two selected connectives *but* and *aber* to trace the transformation pattern from connectives with a weaker signal to connectives with a stronger signal. In general, we assume that explicitation requires more cognitive effort from a translator, whereas implicitation or equivalence do not do so. At the same time, the more ambiguous a connective is, the higher a cognitive effort for its translation. So, ambiguity or the strength of a signal may also have an impact on a translator's behaviour.

Proceedings of 3rd Workshop on Computational Approaches to Discourse, pages 42–47
October 16 – 17, 2022. ©2022 International Conference on Computational Linguistics

The remainder of the paper is organised as follows: in Section 2, we briefly outline the related work. Our methodology is explained in Section 3. We describe the analyses performed in Section 4. In Section 5, we discuss the results and outline our plans for future work.

2 Related Work

Explicitation in translation occurs when a translated text contains new linguistic units not present in the source or more specific linguistic units are used instead of more general units in the source (Klaudy and Károly, 2005, p. 15). Explicitation or implicitation through discourse connectives (as increased or reduced usage of discourse connectives, Olohan and Baker, 2000; Blum-Kulka, 1986), as well as the factors influencing these phenomena, have been analysed in various studies on both human and machine translation (see Shi et al., 2019; Hoek et al., 2015; Zufferey and Cartoni, 2014; Meyer and Webber, 2013).

There are studies showing that explicitation and implicitation may also depend on the type of relation a discourse connective triggers. For instance, cognitively complex relations (e.g. relation of contrast) are not so often left implicit than cognitively simple ones (see Hoek et al., 2017; Blumenthal-Dramé, 2021).

3 Methodology

We use the CRITT translation process database (CRITT TPR-DB, Carl et al., 2016), which has been collected over years and contains a substantial amount of translation process data from numerous translation sessions. The collected data contains features allowing an in-depth assessment of human behavior in translation. We use a part of the data that includes English-German parallel texts. The experiment for this data was set up in such a way that each translator translated every text[1] in one of the three modes: translating from scratch, post-editing and performing monolingual post-editing. Texts were permuted between successive translators, with the intention that for each set of 6 translators, every text would be translated, post-edited and edited 6 times. Each of the 6 source texts is between 110 and 161 words in length and designed in such a way that it fits on one Translog screen (see more details in Carl et al., 2015).

[1]With 32 translators and 6 English source texts in total.

We selected a number of features reflecting translation behaviour, see Table 1. Production pauses reflect the cognitive processes involved in changing attentional states (Schilperoord, 1996): we may assume that if translators start the typing process, they either finalised translation of what they had in mind, or they faced a problem during the writing process. Production pauses can also occur during monitoring, revision and source text reading. Several studies (Kumpulainen, 2015; Lacruz and Shreve, 2014; O'Brien, 2006) have argued that pauses in the flow of keystrokes are indicators of cognitive effort, with longer pauses indicating extended cognitive effort. Pauses are also related to the notion of `first translation response universal` (Carl, 2021), i.e. longer pauses would indicate more entangled activation of the linguistic resources and follow in more challenged and less literal translation (Malmkjær, 2011). The total reading time of the source or the target segments should indicate where the processing effort was located, and whether more attention was drawn to the source or to the target text. Since the database contains 32 translation variants of the English source texts, word translation probabilities are available, too, which have been found to be reverse proportional of cognitive processing effort (see e.g., Carl and Schaeffer, 2017).

4 Analyses

4.1 General cases

First, we extracted the overall connective explicitation in the data at hand without considering explicitation from a weaker to a strong signal. For this, we extracted instances of connectives marked with the part-of-speech label KON on the target side that were aligned to a zero in the source (Query 1 in Table 2). In total, we found 18 cases of this kind of explicitation that we call explicitation insertions. The connective *und*, see example (2), is the most frequently used explicitation insertion (11), followed by *aber* (5), *sowie* (1) and *denn* (1).

(2) *Analysts have warned that prices will increase further still, making it hard for the Bank of England to cut interest rates...*
Analysten haben gewarnt, dass die Preise weiterhin steigen werden und es der Bank of England Probleme bereiten wird, das Zinsniveau zu senken...

We look at the pause (Pause) that precedes the pro-

Feature	Definition
Pause	Typing pause preceding the production unit (i.e. its first keystroke).
TrtS/TrtT	Total reading time in translation unit in the source/target window, refers to the sum total of all fixation durations on a particular area of interest (e.g. token) irrespective of when these occurred during the session.
ProbT	Probability of current translation choice

Table 1: Features available in the CRITT TPR-DB used in the analyses

1	tTokens[((tTokens.PoS==’KON’))& (tTokens.SGroup==’- - -’))]	3	set(sTokens[(sTokens.SToken==’but’)].TGroup)
2	sTokens[((sTokens.PoS==’CC’))& (sTokens.TGroup==’- - -’))]	4	set(sTokens[(sTokens.SToken==’and’)].TGroup)

Table 2: Queries used for the searches in the CRITT TPR-DB

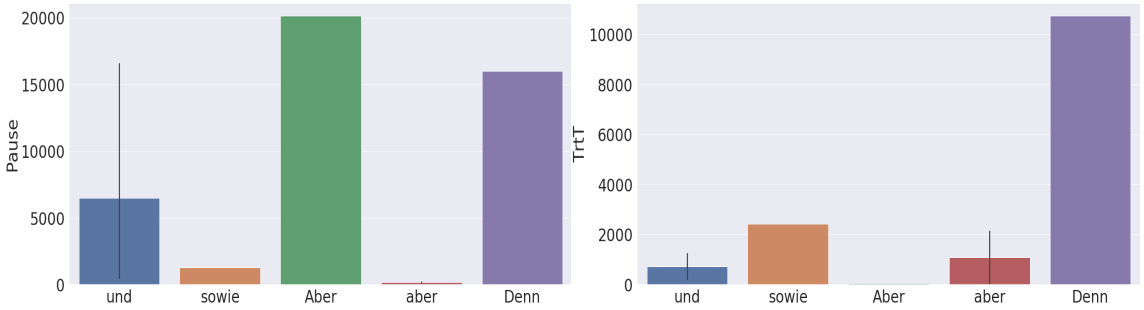

Figure 1: Typing pauses and total reading time on the target of the explicitation connectives

duction of the translation, as well as the total fixation time on the target token (TrtT) for the extracted cases of explicitation, see Figure 1. The longest pauses are observed for the connectives at sentence start. This may indicate processes of translation finalisation of the previous sentence or reading and comprehension activity of the next sentence or phrase. We also observe longer pauses before the production of *und*, which is the most ambiguous connective here (expresses relations of expansion, comparison or time)[2]. This ambiguity causes longer pauses, as activation and selection of linguistic equivalents for more ambiguous items is more challenging. Interestingly, no or very short pauses are observed before the production of the connective *aber*. Explicitation of this connective may require less processing effort because it commonly signals the relation of contrast, a complex relation, and according to the existing studies (see Hoek et al., 2017) is more frequently made explicit as compared to simple relations. At the same time, the total reading times of the explicitation cases with *aber* and *und* are similar.

(3) *Some of the most vulnerable countries of the world have contributed the least to climate change, but are bearing the brunt of it. Einige der Länder, die den Klimawandel am härtesten zu spüren bekommen, haben nur sehr wenig dazu beigetragen.*

Next, we extracted cases of implicitation, i.e. when a connective in the source (marked as CC) is left out, see Query 2 in Table 2 above. The query extracted 11 cases with the connectives *and* (9) and *but* (2). However, manual validation revealed that the query results contained noise does and only one case of implicitation, as illustrated in example (3), where the connective *but* was left out in the German translation.

4.2 Specific connectives

Then, we extracted all cases of translations of the discourse connective *but* (Query 3 in Table 2 above). The results of the query show that our data contains translations with *aber*, *doch*, *jedoch* and *obwohl*, as well as implicitation (the connective was left out). While we consider translations with *aber* as an equivalent, translations with *doch*, *jedoch*, *obwohl* are explicitation cases, as these connectives trigger one type of relations only and hence, convey a stronger signal than *but*[3]. Production pauses and the total reading time in the source (*but*) and the target tokens (*aber*, *doch*, *jedoch*, *obwohl*), as well as the probability of translations are visualised in Figure 2. As expected, implicita-

[2]see Connective-Lex, the web-based multilingual lexical resource (Stede et al., 2019).

[3]The ambiguity of was verified with the help of Connective-Lex.

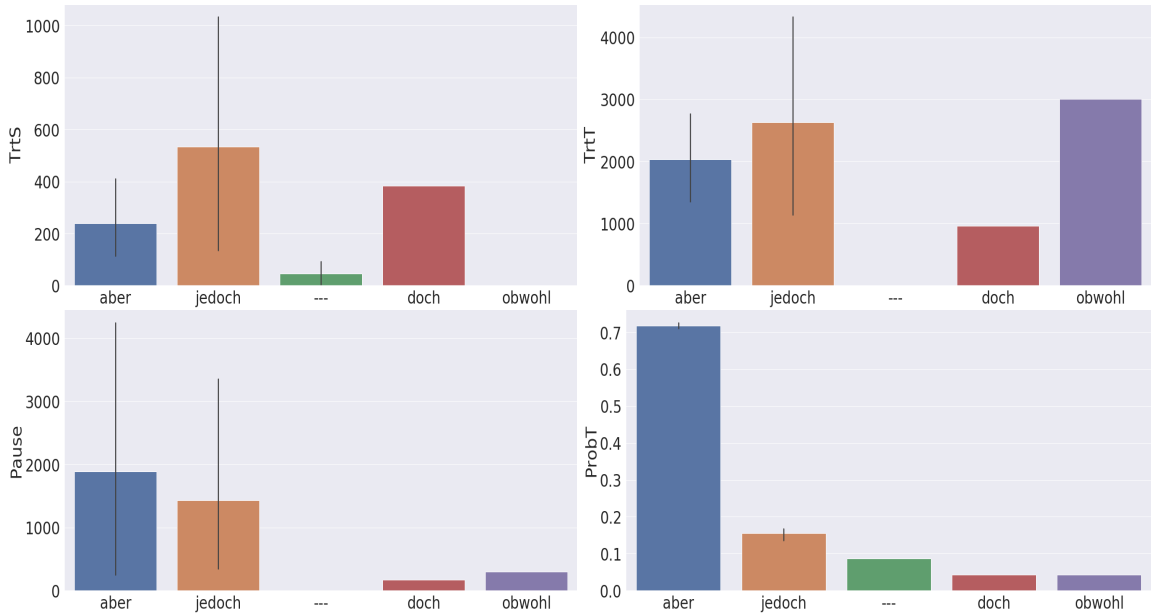

Figure 2: Total reading time, typing pauses and probability of translations of *but*

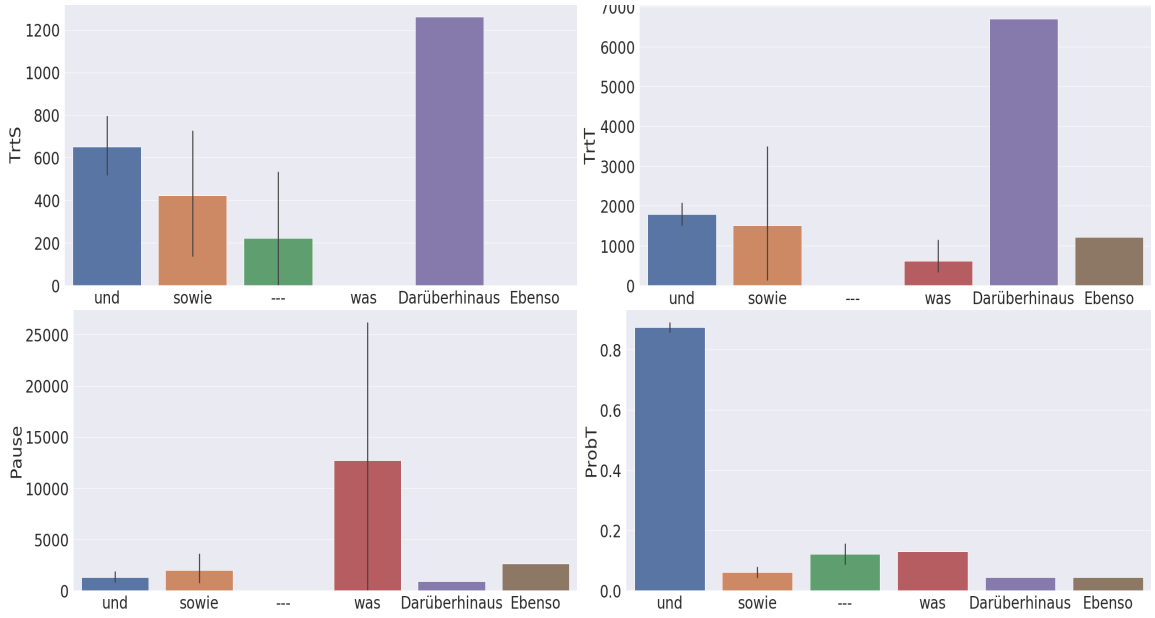

Figure 3: Total reading time, typing pauses and probability of translations of *and*

tion requires the lest effort, which is reflected in no pauses and short reading time. However, for the equivalence translation with *aber*, we observe the longest production pauses, which is against our expectations. For explicitation with *jedoch*, we observe longer pauses and longer reading times. Generally, this discourse connective is less frequent in the data and may need additional time for mental activation. Explicitation with *obwohl* has the longest reading time in the target and no reading time on the source. Both *obwohl* and *doch* have the shortest pauses. Given the source connective *but*, the connective *aber* has the highest translation probability, followed by *jedoch*, omission, *doch* and *obwohl*. The graphs reveal that higher translation probability generally causes longer production pauses.

Production pauses and the total reading time of the source *and* and the target tokens (*und, sowie, was, Darüberhinaus, Ebenso*), as well as omissions were extracted with Query 4 in Table 2 and are visualised in Figure 3.

The longest production pause is observed for *was*, which is not a connective but rather a pronoun

referring to the previous clause, see example (4).

(4) *Incentives must be offered to encourage developing countries to go the extra green mile and implement clean technologies, and could also help minimise...*
Es werden daher Anreize angeboten, Entwicklungsländer zu fördern, um ihnen zusätzlich grüne Standards zu ermöglichen und saubere Technologien zu implementieren, was auch zur Minimierung...

This case is also indicated by low reading time on the target (and no reading time on the source). Similarly to translations of *but*, no pause and the shortest reading time are observed when *and* is left out, confirming that implicitation does not require a high cognitive effort. This indicates that little/no cognitive effort is required for implicitation of comparison and expansion relations triggered by *but* and *and*. Equivalence translation is featured by a very short pause (different to what we observed in the case of *but*), but high reading time. Interestingly, there is an opposite tendency here in terms of the source vs. target reading time: in equivalence translation of *but*, it was shorter on the source (also generally common in the process of translation), whereas for *and* translated as *und*, it is longer on the source. This could be due to the greater ambiguity of *and*, if compared to *but*. The highest reading time, but short pauses, are reported for the explicitation with *Darüberhinaus*. The longest pause is observed for explicitation with *was*. The equivalent connective *und* has the highest translation probability given the source *and*. Here, translation probability does not necessarily causes longer pauses and hence greater cognitive load, which is different to the cases with *but*.

5 Summary and Discussion

We attempted to analyse explicitation and implicitaion phenomena of discourse connectives in English-German translations using the parallel data from the CRITT TPR-DB. Our results show that while implicitation requires low cognitive effort, it is not necessarily so for an equivalent translation. This may depend on the connective as indicated by the differences in pauses observed. This may also be dependent on the strength of its signal and the type of relation this connective conveys. Explicitation generally causes a higher effort in the analysed cases, which are however quite few. In the future, we would like to analyse more instances of explicitation and implicitation for more connectives and include data originating from different genres, as there could be variation in processing discourse connectives across different contexts. Moreover, we also intend to analyse differences in the cognitive processing of connectives depending on the processes of translation – if a translation was produced from scratch or if it was post-edited from a machine-translated output. This will provide some insights on how human translators are impacted by discourse-related issues in machine-translated texts. Moreover, translation process data provides explanations about problems human translators face. These may correlate with the difficulties in machine translation. A better understanding of problems in human translation may also help to improve machine translation.

References

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2020. Translation artifacts in cross-lingual transfer learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7674–7684, Online. Association for Computational Linguistics.

Fatemeh Torabi Asr and Vera Demberg. 2012. 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. The COLING 2012 Organizing Committee.

Mona Baker. 1993. Corpus linguistics and translation studies: Implications and applications. In G. Francis Baker M. and E. Tognini-Bonelli, editors, *Text and Technology: in Honour of John Sinclair*, pages 233–250. Benjamins, Amsterdam.

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

Alice Blumenthal-Dramé. 2021. The online processing of causal and concessive relations: Comparing native speakers of English and German. *Discourse Processes*, 0(0):1–20.

Michael Carl. 2021. Micro units and the first translational response universal. In Michael Carl, editor, *Explorations in Empirical Translation Process Research*, Machine Translation: Technologies and Applications, pages 233–257. Springer.

Michael Carl, Silke Gutermuth, and Silvia Hansen-Schirra. 2015. Post-editing machine translation: Efficiency, strategies, and revision processes in professional translation settings. In *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting*, volume 115 of *Benjamins Translation Library*. John Benjamins.

Michael Carl and Moritz Schaeffer. 2017. Sketch of a noisy channel model for the translation process. In Silvia Hansen-Schirra, Oliver Czulo, and Sascha Hofmann, editors, *Empirical Modelling of Translation and Interpreting*, volume 7 of *Translation and Multilingual Natural Language Processing*, pages 71–116. Language Science Press, Berlin.

Michael Carl, Moritz Schaeffer, and Srinivas Bangalore. 2016. The CRITT translation process research database. In Michael Carl, Srinivas Bangalore, and Moritz Schaeffer, editors, *New Directions in Empirical Translation Process Research*, New Frontiers in Translation Studies, pages 13–54. Springer Science+Business Media, Germany.

Ludivine Crible. 2020. Weak and strong discourse markers in speech, chat and writing: Do signals compensate for ambiguity in explicit relations? *Discourse Processes*, 57(9):793–807.

Koel Dutta Chowdhury, Cristina España-Bonet, and Josef van Genabith. 2020. Understanding translationese in multi-view embedding spaces. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6056–6062, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Martin Gellerstam. 1986. Translationese in Swedish novels translated from English. In L. Wollin and H. Lindquist, editors, *Translation Studies in Scandinavia*, pages 88–95. CWK Gleerup, Lund.

Yvette Graham, Barry Haddow, and Philipp Koehn. 2020. Statistical power and translationese in machine translation evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 72–81, Online. Association for Computational Linguistics.

Jet Hoek, Jacqueline Evers-Vermeul, and Ted J.M. Sanders. 2015. 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. Association for Computational Linguistics.

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

Kinga Klaudy and Krisztina Károly. 2005. Implicitation in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures*, 6:13–28.

M. Kumpulainen. 2015. On the operationalisation of 'pauses' in translation process research. *The International Journal for Translation & Interpreting Research*, 7:47–58.

I. Lacruz and GM. Shreve. 2014. Pauses and cognitive effort in post-editing. In O'Brien S., L. Balling, M. Carl, M. Simard, and L. Specia, editors, *Post-editing: processes, technology and applications*, pages 246–272. Cambridge Scholars Publishing, Newcastle upon Tyne.

Kirsten Malmkjær. 2011. Translation universals. In Kirsten Malmkjær and Kevin Windle, editors, *The Oxford handbook of translation studies*, pages 83–93. Oxford University Press, Oxford.

Thomas Meyer and Bonnie Webber. 2013. Implicitation of discourse connectives in (machine) translation. In *Proceedings of the Workshop on Discourse in Machine Translation*, pages 19–26, Sofia, Bulgaria. Association for Computational Linguistics.

Sharon O'Brien. 2006. Pauses as indicators of cognitive effort in post-editing machine translation output:. *Across Languages and Cultures*, 7(1):1 – 21.

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

Joost Schilperoord. 1996. The distribution of pause time in written text production. In Gert Rijlaarsdam, Huub van den Bergh, and Michel Couzijn, editors, *Theories, models and methodology in writing research*, pages 21–35. Amsterdam University Press.

Weijia Shi, Muhao Chen, Yingtao Tian, and Kai-Wei Chang. 2019. 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. Association for Computational Linguistics.

Manfred Stede, Tatjana Scheffler, and Amália Mendes. 2019. Connective-lex: A web-based multilingual lexical resource for connectives. *Discours*, 24.

Gideon Toury. 1995. *Descriptive Translation Studies – and Beyond*. John Benjamins, Amsterdam.

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