Translation in Transition 6

22–23 September 2022
Charles University, Prague

Book of Abstracts

Table of Contents

3

Plenaries

Silvia Bernardini

Translation and where to find it: Corpus methods for 21st century translation research

This is a time of both anxiety and excitement for the translation sector. Global connectedness and dramatic improvements in artificial intelligence mean that prototypical professional translation, or *translation as we knew* it, is likely to become marginal to society in the short to medium term. It no doubt will be replaced, however, by a range of activities characterized by varying degrees of professionalism, flexible multilingualism and human-computer collaboration that will make *translation broadly conceived* even more central. This opens up exciting perspectives for empirical research on translation, but arguably calls for reflection on our established research paradigms, from the very basic issue of defining and delimiting our object of study, through to the range of methods we use to investigate it. To make this discussion more concrete, I will refer to two very different scenarios: one concerning machine- and learner-translated texts, the other translation in multilingual news production. Both case studies point to the centrality of parallel corpora as a research infrastructure, but at the same time signal the need for more flexibility in how we conceive of well-established translation and corpus categories and assumptions.

Gert De Sutter

Taking stock of corpus-based translation studies, and some suggestions for further improvement

Corpus-based translation studies has come a long way since it started in the 1990's: (i) by scrutinizing patterns in parallel and comparable corpora, it has found abundant and reliable evidence that language use produced in a translational context (often: published interlingual translations by expert translators, with English as source or target language) differs significantly from language use produced in non-translational contexts; (ii) due to major methodological and analytical improvements, often driven by innovations in corpus linguistics, our view on *when*, *how and to what* extent translated language differs from non-translated language has become increasingly more accurate; (iii) finally, in more recent years, these descriptive and methodological advances are accompanied by more solid theoretical reasoning, providing (tentative) answers on *why* translated language use differs from non-translated language use. In this talk, I want to take stock of these achievements by focussing on the growing convergence of description, method and theory in corpus-based translation studies and by exploring (some of) the limits that current corpus research has to do deal with. This will lead me to suggest at least three further (methodological) improvements to the field:

- compiling new-generation parallel corpora which contain more and more carefully checked metadata on the translators, the translation project and the source/target texts incorporated in the corpus.
- adopting new procedures in the statistical analysis of data patterns in corpora (a.o. multilevel modelling and naive discriminative learning)
- adding hitherto underexplored predictor variables.

I will argue that these improvements will further contribute to the increasing convergence of description, method and theory in corpus-based research designs, thereby fundamentally enhancing our understanding of language products of and language production during translation.

Full papers

Vahram Atayan
Institute for Translation and Interpreting, Heidelberg University
atayan@uni-heidelberg.de

Bogdan Babych
Institute for Translation and Interpreting, Heidelberg University
bogdan.babych@iued.uni-heidelberg.de

Translating argumentation: distributional semantic analysis of argumentation resources in parallel corpora

1. Research question and context

The characterization of linguistic resources used in argumentation has been an important challenge for contrastive linguistics, qualitative and corpus-based translation studies (e.g. Atayan 2007). In particular, different languages build argumentation structures with different inventories of lexical, morphosyntactic, and discursive means, as well as usage patterns, which involve interaction across various linguistic levels and paradigmatic sub-systems of a language (argumentative connectors, evaluative/connoted lexicon, meta-argumentative constructions, etc.). Also syntagmatically the argumentation patterns may span non-local context, extending through several sentences and larger discourse units. However, it is difficult to automatically identify and align multiword argumentation patterns in multilingual corpora with sufficient accuracy. This seriously limits the applicability of standard corpus-based methods, tools and annotation resources for their study, since most traditional approaches primarily target phenomena in the local context of corpus searches (e.g., morphosyntactic or lexical patterns within a window of a few words). Specifically, while there have been several corpus-based studies of systematic differences between original texts and translations (so-called 'translationese') on the levels of the general lexicon, modal markers, morphosyntactic patterns, indirect equivalents (e.g., Babych et al., 2007; House, 2011; Hoey, 2011; Kranich and Gast, 2015; Gast, 2022), the range of such studies for argumentation patterns across languages has been limited.

2. Methods
Our paper addresses this methodological gap in contrastive corpus-based analysis of argumentation, namely we suggest a number of new distributional semantic properties of argumentation resources, which allow us to quantify differences in their usage across languages and genres or in original and translation corpora within the same language. In the paper we report the results of our experiments on evaluating relevant distributional parameters of three types of argumentation patterns in multilingual and translation context: (a) meta-argumentative words, like '*disagree*' or '*reason*'; (b) *key notions* of a given discourse, e.g, '*Commission*' for EU parliament proceedings or terms like '*venous*' for a medical corpus; (c) e*valuatives/connotated* lexicon, like 'dangerous' or 'progress'. Lexical items of these three categories have been manually annotated in two different selections of ca. 1000 word types from each of our corpora: (1) The Europarl corpus of parliamentary proceedings (Koehn, 2005), where we selected original texts that have been authored in (1a) German and (1b) English, as well as (1c) English texts translated from German. (2) The CORD-19 English monolingual corpus of medical research articles about Covid-19 (Wang et al., 2020). Both corpora are POS-tagged and lemmatized using TreeTagger (Schmid, 2013)..

3. Results
3.1. Knee of MI-ranked collocate lists as a distributional measure for argumentation lexicon
Argumentation is a phenomenon of communicative discourse (van Eemeren, 2018; van Eemeren et al., 2019, 5), so an important question to address is identification of the suitable collocation spans for the analysis of the argumentative lexicon. We suggest using the quantity of relevant collocates as a measure for identifying the most suitable spans for the different types of argumentative words. We consider as relevant the most informative collocates with an above-average contribution to the overall Mutual Information (*MI*, e.g., Evert, 2008) of the whole collocate set. This number is calculated as the knee of the discrete curve of decreasing MI values (Satopää et al., 2011), which represents the best balance for inherent trade-offs between stronger collocates and rather accidental correlations.

The dynamics of knee change across different spans indicates that meta-argumentative lexicon has distinct distributional characteristics in comparison with the general lexicon and other types of argumentation resources,

(evaluatives, key notions, etc.), which could indicate that these linguistic constructions find 'islands of consistency' both in the local and more distant context within the discourse (e.g., Enghels and Sansiñena, 2021), possibly forming a part of `coordinated constructions' (Fillmore et al., 1988). In the translation context, meta-argumentatives are also much stronger influenced by the target language in comparison to other types of argumentation lexicon, possibly because of greater asymmetry in the syntactic structure of English and German clauses, which could have a particular impact on meta-argumentative verbs.

3.2. Word-vector centroid-based prediction for argumentation lexicon
In our last experiment we use the word embedding models (Mikolov et al., 2013) to identify candidates for argumentatively-relevant lexical items, and to evaluate the potential improvement in precision. The lists of candidates are generated as closest words to the respective centroids (Brokos et al., 2016, 114) of annotated *key notions, evaluatives* and *meta-argumentative* elements. We find a higher percentage of items of a given argumentative category with respect to the initial annotation for 22 of 24 centroids, with an improvement up to 6,6 times for some categories. Our method also provides a general improvement for the selection of argumentatively-relevant lexical items from the whole set of potential candidates.

4. Conclusions
Our results indicate that distributional characteristics of argumentation resources can be modelled with the proposed discourse-level metrics, such as the average size of knee of ranked collocation lists and vector centroids for word embeddings of argumentation lexicon. Specifically, these methods reduce noise in larger collocation spans, which are needed for capturing distributional properties beyond local context.

References

Atayan, V. (2007). Argumentationsstrukturen – ein Äquivalenzparameter bei der Übersetzung. *Multiperspektivische Fragestellungen der Translation in der Romania. Hommage an Wolfram Wilss zu seinem, 80*, 61–93

Babych, B., Hartley, A., Sharoff, S., & Mudraya, O. (2007). Assisting translators in indirect lexical transfer. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 136–143.

Brokos, G. I., Malakasiotis, P., & Androutsopoulos, I. (2016). Using Centroids of Word Embeddings and Word Mover's Distance for Biomedical Document Retrieval in Question Answering. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, pages 114–118, Berlin, Germany, Aug. 2016. Association for Computational Linguistics. doi: 10.18653/v1/W16-2915. URL https://aclanthology.org/W16-2915.

Enghels, R., & Sansiñena, M. S. (2021). Discourse-level phenomena in construction grammars. *Constructions and Frames, 13*(1), 3–20.

Evert, S. (2008). Corpora and collocations. In: A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook* (σσ. 1212–1248). URL: http://purl.org/stefan.evert/PUB/Evert2007HSK_extended_manuscript.pdf

Fillmore, C. J., Kay, P., & O'connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 501–538.

Gast, V. (2022). Comparing Annotation Types and n-Gram Sizes. In: O. Schützler & J. Schlüter (Eds.), *Data and methods in corpus linguistics: Comparative approaches* (pp. 323–352). Cambridge University Press.

Hoey, M. (2011). Lexical priming and translation. In: *Kruger, Alet, Kim Wallmach & Jeremy Munday (eds.) Corpus-based translation studies: Research and applications*, 153–168.

House, J. (2011). Using translation and parallel text corpora to investigate the influence of global English on textual norms in other languages. In: *Kruger, Alet, Kim Wallmach & Jeremy Munday (eds.) Corpus-based translation studies: Research and applications*, 187–208.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. *Proceedings of machine translation summit x: papers*, 79–86.

Kranich, S., & Gast, V. (2015). Explicitness of epistemic modal marking: Recent changes in British and American English. *Thinking Modality: English and Contrastive Studies of Modality. Newcastle upon Tyne: Cambridge Scholars Publishing*, 3–22.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*. URL: http://arxiv.org/abs/1301.3781

Satopää, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a" kneedle" in a haystack: Detecting knee points in system behavior. *2011 31st inter-*

national conference on distributed computing systems workshops, 166–171. IEEE.

Schmid, H. (2013). Probabilistic part-of-speech tagging using decision trees. *New methods in language processing*, 154-164.

van Eemeren, F. H. (2018). *Argumentation theory: A pragma-dialectical perspective.* Springer.

van Eemeren, F. H., Grootendorst, R., & Kruiger, T. (2019). Handbook of argumentation theory. In: *Handbook of Argumentation Theory*. De Gruyter Mouton.

Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Burdick, D., … Kohlmeier, S. (2020, July). CORD-19: The COVID-19 Open Research Dataset. *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020.* URL: https://aclanthology.org/2020.nlpcovid19-acl.1

••••

Cecile Cuchet
Université Lyon 2, CeRLA
cecile.cuchet1@univ-lyon2.fr

Pascaline Dury
Université Lyon 2, CeRLA
pascaline.dury@univ-lyon2.fr

The use of Knowledge Rich Diachronic Information to translate innovative and emergent specialised fields

The proposed paper aims to describe and evaluate the use of diachronic information to help translation students gain a better understanding of concepts and terms in highly complex scientific domains, particularly in emergent and/or innovative scientific fields, such as nanomedicine. A number of research works show the value of integrating diachronic information into specialised language courses (Celotti and Musacchio, 2004 ; Kawaguchi et al., 2011 or Heremans and Cuyckens, 2012), but none of them has yet explored the use of short diachronic specialised corpora for teaching specialised translation. The paper is based on the assumption that the diachronic perspective provides students with cognitive and terminological information that helps them become familiar with the conceptual organisation of a specialised field and its recent evolution. We postulate that this type of information is essential when translating emergent specialised domains and/or domains in which innovation and progress are constant, such as nanomedicine, and for which lexicographical and terminological resources are therefore scarce or insufficiently updated.

This hypothesis has been tested in practice over 3 non-consecutive years (2016-2017, 2018- 2019, 2021-2022) with the students following the professional translation courses of the Master 2 in medical translation at the University Lyon 2. In a first part of the experiment, the students had to compile and analyze a comparable bilingual corpora (English and French), covering a short period of time (17 years, from 2004 to 2021), in the field of nanomedicine applied to cancer treatment, in order to prepare a glossary of terms for their translations. The second part of the experiment consisted in evaluat-

ing which diachronic information extracted from the corpus had been the most useful for the students to understand the terminology of the field and to grasp the main conceptual changes having taken recently place in the domain.

Hence, a first list of what we call knowledge-rich diachronic information (KRDI) was established . It includes, for instance, information on the different generations of terms, and on the sometimes very rapid replacement (in less than a year) of one generation by another (e.g. from *carbon nanotubes* to *single-walled carbon nanotubes* to *multi-walled carbon nanotubes* up to the most recent *f-engineered carbon nanotubes*); information on families of terms (e.g. the appearance of term families referring to the concept of networks, such as *microfluidic network* or *functional vessel network* and to the notion of complex assembly, *siRNA transfection complex*, *DNA duplexe*); and information on the synonymic variation characterising the appearance of new terms (*small interfering (si)RNAS* or *small interference RNA (siRNA)* or *small interfering RNA (siRNA))*. The presentation will therefore focus on the advantages of using such diachronic corpora in the training of translation students, but will also discuss its limitations (e.g. diachronic information is not very useful to help understand the use of the very complex abbreviations used in the field).

References

Celotti, N. et M. T. Musacchio. (2004). Un regard diachronique en didactique des langues de spécialité. *Ela. Etudes de linguistique appliquée*, 3 (135), 263-270.

Heremans, K. et H. Cuyckens.(2012). DIACHRONEX: Corpus-based exercises for English diachronic linguistics, *ICAME Journal,* 36, 67-93.

Kawaguchi, Y., Minegishi, M. et W. Viereck. (2011). *Corpus-based Analysis and Diachronic Linguistics.* Amsterdam, Philadelphie : John Benjamins Publishing Company.

••••

Ella Diels
University of Antwerp
ella.diels@uantwerpen.be

Jim J. J. Ureel
University of Antwerp
jim.ureel@uantwerpen.be

Isabelle S. Robert
University of Antwerp
isabelle.robert@uantwerpen.be

Carola Strobl
University of Antwerp
carola.strobl@uantwerpen.be

Using corpus-focused instruction in translation (revision) training to develop stylistic translation revision competence: A mixed-methods study

In this paper, we focus on the use of corpora in translation training. More specifically, we examine the effects of corpus-focused instruction (CFI) on the development of sensitivity to stylistic translation revision competence (TRC). In a pilot study, we found that translation trainees (L1 Dutch–L2 English) struggle with accommodating language to formal English communicative contexts (Ureel et al., 2022). In this study, we build on these results and examine if CFI can be a possible solution to the pedagogical challenges inherent in developing (L2) sociolinguistic competence. CFI is a form of data-driven learning, a language learning method which 'confront[s] the learner as directly as possible with the data' (Johns, 2002, p. 108). Other studies demonstrated the positive effects of corpora in translation training (e.g., Beeby et al., 2009; Boulton, 2012; Zanettin et al., 2003). We investigate if and how CFI can improve translation trainees' ability to detect and correct instances of stylistic inappropriateness in English translations of Dutch academic texts. We compared two kinds of CFI – 'do-it-yourself' (DIY) corpora and prepared (i.e. existing) corpora – to answer two research questions: (1) Can CFI help translation trainees to make more informed style-related decisions during translation revision? and (2) Which form of CFI triggers the highest success

rate? To answer these research questions, we used an explanatory sequential design. In Phase 1, we collected quantitative data using a (quasi-)experimental crossed-conditions design. We assigned 32 graduate translation trainees (L1 Dutch–L2 English) to two groups. For the CFI, we developed teaching units in which we introduced Sketch Engine as a translation (revision) tool and explained how corpora can be used to solve stylistic translation (revision) problems. In the teaching units, Group 1 focused on how to build and use DIY corpora and Group 2 was made more aware of register variation in (multilingual) corpora. All the translation trainees participated in four stylistic revision tests. We first organised a pretest to assess the participants' levels pre-experimentally. After the pretest, Group 1 received CFI on DIY corpora. Next, we organised the intermediate test. After the intermediate test, Group 2 received CFI on prepared corpora. The final treatment phase was followed by an immediate posttest and a delayed posttest (6 weeks after the immediate posttest). To analyse the quantitative data, we examined the revision products (i.e. the revisions made by the participants) and the revision processes. In Phase 2, we collected qualitative data by conducting semi-structured interviews with stimulated recall. We interviewed four participants from Group 1 and five participants from Group 2. The quantitative data show that the CFI did not have a statistically significant effect on the participants' performance on the stylistic revision tests. Although Groups 1 and 2 followed the CFI at different points in time, both groups experienced the same development. That is, Group 1 did not perform statistically significantly better than Group 2 on the intermediate test (i.e., the test following the CFI on DIY corpora and preceding the CFI on prepared corpora). Moreover, Group 2 did not perform statistically significantly better on the immediate posttest (i.e., the test following the CFI on prepared corpora) than on the intermediate test. In contrast, the CFI did affect the participants' revision processes as both groups started to use corpora effectively after the CFI. Our qualitative data confirm the quantitative findings, as most interviewees explained that they had continued to use Sketch Engine for other translation (revision) and writing assignments (e.g., other university courses, internships). In conclusion, the CFI did not have a statistically significant effect on the participants' performance on the stylistic revision tests, but the data do show that corpora can be used as effective translation revision tools. More research is required to find out (1) why the CFI did not affect the participants' performance on the stylistic revision tests and (2) how the CFI could be adjusted to improve success rates in the future.

References

Beeby, A., Rodríguez Inés, P., & Sánchez-Gijón, P. (Eds.). (2009). *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate.* Amsterdam: John Benjamins.

Boulton, A. (2012). Beyond concordancing: Multiple affordances of corpora in university language degrees. *Elsevier Procedia: Social and Behavioral Sciences,* 34, 33-38.

Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Linguistics.* Amsterdam: Rodopi, 107-117.

Ureel, J. J. J., Diels, E., Robert, I. S., & Schrijver, I. (2022). The development of L2 sociolinguistic competence in translation trainees: An ccommodation-based longitudinal study into the acquisition of sensitivity to grammatical (in)formality in English. *The Translator and Interpreter Trainer*, 16 (1), 78-95

Zanettin, F., Bernardini, S., & Stewart, D. (Eds.). (2003). *Corpora in Translator Education.* Manchester: St Jerome.

••••

Anna Espunya
Universitat Pompeu Fabra, Spain
anna.espunya@upf.edu

Measuring shifts in abstractness with cognitive psychology instruments

It has been claimed that in formulating a text in English, linguistic realization mirrors concrete reality to a greater extent than in other languages, such as French (Vinay and Darbelnet 1995: 51) and Spanish (Vázquez Ayora 1977: 82- 84). Details of physical reality contribute to this "concrete level of expression", while abstract and generic words (re)present content at an "abstract level of expression." Translation from Spanish into English, according to individual case studies, involves the choice of "concrete language" (e.g. Munday 2008: 219-220) on the English translation of *One Hundred Years of Solitude*). Such observations have not been empirically tested, probably due to the difficulty in operationalizing the concrete-abstract dichotomy, which requires a combination of formal criteria and semantic analysis.

Among the formal criteria, morphological structure –such as suffixation– is one of the parameters predicting abstract words (Reilly and Kean, 2007; Strick Lievers, Bolognesi and Winter, 2021). In Author (2020) instances of Spanish deadjectival nouns formed by suffixation with -*idad* (e.g. *amabilidad* 'kindness', related to *amable* 'kind') and their correspondences in English translations are obtained from a parallel corpus of two contemporary fiction works and their respective translations by different translators. The subset of -*idad* instances with an evaluative function, where the deadjectival noun denotes a quality of an entity anchored in the fictional world, were selected and manually tagged for translation shifts in *conceptual* and *grammatical* abstractness. Conceptual shifts were identified as linguistic choices involving semantic variations toward physical and/or perceptible qualities. Shifts in grammatical abstractness were defined as those reverting reification through the choice of adjectives, adverbs and verbs instead of deadjectival nouns. Often this involves modifying the original semantic structure, shifting the object of the evaluation and potentially altering authorial intention of foregrounding the qualities as opposed to entities (metaphor).

Assuming as the null hypothesis that no decrease in abstractness should be observed in English translations of Spanish texts, the study found statistically significant decreases in conceptual abstractness in both English translations: first translation ($\chi 2$ (1, N=46) = 7.485, P= .0062); second translation ($\chi 2$ (1, N=39)= 7.569, P=.0059). Decrease in grammatical abstractness was even more statistically significant: first translation ($\chi 2$ (1, N=46) = 17.722, P<0.0001) and second translation ($\chi 2$ (1, N=39) =23.130, P<0.0001).

One of the potential shortcomings of that study is its manual annotation procedure, particularly for the conceptual abstractness variable. In order to correct this, a new approach has been tested based on wordlists rated for conceptual concreteness that are widely used in cognitive psychology studies, namely Brysbaert et al. (2014) for English and Guasch et al. (2016) for Spanish. Such word ratings are obtained in large-scale informant studies of lexical items through carefully designed questionnaires and offer the advantage of curated interrater agreement. Each -*idad* word in the previous sample was assigned the concreteness rating provided by the Spanish wordlist and paired with the score provided by the English wordlist to its translation correspondence.

An independent-samples t-test was conducted to compare mean conceptual concreteness scores in each Spanish text (SPT) and its English translation (ENT). The comparisons yielded mixed results. For SPT 1- ENT1, the difference in concreteness is not significant [t (30)=2.894, p = .6559; SPT1 (M= 2.19, SD=0.76), ENT1 (M=2.30, SD=0.60)]. For SPT2 – ENT2, the difference was significant [t (30)=2.894, p = .007; SPT2 (M=2.06, SD=0.42) and ENT2 (M=2.62, SD=0.65)]. Translation in the Spanish-English direction does not always involve the choice of more concrete lexical items. Differences might be attributed to the strategies used by individual translators, a hypothesis that requires testing a larger and more translator-diverse corpus.

The use of rated wordlists is not problem-free. On a practical level, the Spanish list is much shorter than the English one, an obstacle to pairing all possible ST-TT sequences. On the methodological level, lists are language-dependent in many senses (Strick Lievers et al. 2021: 664). For example, a cognate can be rated as more abstract in one language than in another (see Sp. *vitalitad* vs En. *vitality*, rated 2.36 vs 1.78 on a five-point scale). This, however, is to be expected if speakers of different languages have different linguistic realization preferences, i.e. Spanish raters consider *vitalidad* more concrete than English raters consider *vitality*. Also, following Strick Lievers

et al. (2021) and references therein, it may be the case that raters are sensitive to the linguistic properties of lexical items they are asked to rate, such as word class, suffix types, etymology, etc. Further avenues for research include a more detailed study of the translations to explore the effect of word class choice and the availability of cognates.

The methodology is applicable to research about other genres characterized by the use of abstract lexicon.

References

Author. (2020). Reduced abstractness in Spanish-English translation: the case of property-denoting nouns. *Meta: Translators' Journal* 65 (2).

Brysbaert, Marc, Amy Beth Warriner & Victor Kuperman. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46 (3), 904–911.

Guasch, Marc, Pilar Ferré and Isabel Fraga. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods,* 48 (4), 1358–1369.

MedCalc Software Ltd. Comparison of proportions calculator. https://www.medcalc.org/calc/comparison_of_proportions.php (Version 20.027; accessed January 28, 2022)

MedCalc Software Ltd. Comparison of means calculator. https://www.medcalc.org/calc/comparison_of_means.php (Version 20.027; accessed February 3, 2022)

Munday, Jeremy. (2008). *Style and Ideology in Translation. Latin American Writing in English*. New York/London: Routledge.

Reilly, Jamie and Jacob Kean. (2007). Formal distinctiveness of high- and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science*, 31 (1), 157–168.

Strick Lievers, F., Bolognesi, Marianna, and Bodo Winter. (2021). *Cognitive Linguistics*, 32 (4), 641-670.

••••

Lenka Fárová
Faculty of Arts, Charles University
lenka.farova@ff.cuni.cz

Translating unique items: Finnish verbs of sufficiency and their Czech counterparts

The notion of 'unique items' (Tirkkonen-Condit 2002, 2004), i.e., linguistic elements that are difficult to translate, has been discussed in translation studies for some decades. These elements can be lexical, syntactic, or phrasal, and they lack straightforward counterparts in the target language. Unique items have been so far studied mainly from the point of translation, and it has been suggested that their translations tend to under-represent target-language-specific linguistic features while they over-represent features that have straightforward equivalents which are frequent in the source text (see e. g. Eskola 2004, Tirkkonen-Condit 2004, cf. also Chlumská 2017).

This paper aims to explore specifically how are unique items translated from Finnish into Czech. The working hypothesis assumes that lexical items that lack a clear equivalent are translated by a broad range of expressions which may depend on the genre of the text as well as on the personal style of the translator. Existing bilingual dictionaries can also affect the translation. One type of the often-mentioned unique items in Finnish are the verbs of possibility and sufficiency, i.e., verbs that have as part of their semantics a notion of their having to be a sufficient amount of some entity (e.g., time, energy) at disposal for an action to be possible (Johansson & Nordrum 2021: 67). These verbs do not have, in many Indo-European languages including Czech, monomorphemic or straightforward counterparts, e.g., *jaksaa* ´it is possible for someone to do something because there is a sufficient amount of energy available to perform the task´. Other verbs from this semantic group refer to boldness to overcome embarrassment (*kehdata*), courage (*uskaltaa*), time (*ehtiä*) etc. (see Flint 1980: 61). Flint (1980) lists more than forty such verbs from which only the twelve most frequent are examined in this study: *ehtiä*, *jaksaa*, *riittää*, *uskaltaa*, *kelvata*, *mahtua*, *viitsiä*, *kehdata*, *viihtyä*, *malttaa*, *rohjeta* and *joutaa*.

The aim of this study is not to discuss the type of modality of the verbs (cf. Kangasniemi 1992: 44-49), it is rather to show the broad range of translation

solutions. The heterogeneity of translations is already visible in the entries in the existing Finnish-Czech dictionaries (Lindroos 1984, Lingea 2016). This corpus-based study thus aims to show trends in these translation solutions with the variables of genre, author, and translator in mind. The material was extracted from the parallel corpus InterCorp v13-ud Finnish, subcorpus of Finnish prose and its translation into Czech. The subcorpus consists of 2,000,830 tokens represented by 29 books of different genres (historical, detective, humoristic and Finnish weird novels (Sinisalo 2011), thrillers as well as contemporary prose) written by 8 female and 9 male Finnish authors and translated by 6 different translators. The absolute frequency of the analysed verbs ranges from 878 (*ehtiä*) to 27 (*joutaa*).

The preliminary results show the distribution of the verbs varies across different genres and with different authors, e. g., verb *jaksaa* (492.51 imp) is preferred by the contemporary author Kari Hotakainen, while it is used substantially less (95.29 ipm) by Mika Waltari (1908-1979). The analysis of the translation counterparts has shown that, perhaps unsurprisingly, these verbs are translated with many different equivalents, most of which are not one-word, and that the meaning of the verb is often rendered on the syntactic level rather than on the lexical level. Negative forms of the verbs seem to affect the choice of translation equivalents, too. Some of the expressions have already reached lexicalized forms of translation which are then used by majority of the translators, for example, negative imperative form *Älä viitsi/viitti.*, of the verb *viitsiä,* which refers to a sufficient amount of mental energy, occurs, is most frequently translated as *Nech toho.* [´Stop/leave it.´]).

References

Eskola, S. (2004). Untypical frequencies in translated language. In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: do they exist?* Amsterdam-Philadelphia: John Benjamins, 83-99.

Flint, A. (1980). *Semantic Structure in the Finnish Lexicon: Verbs of Possibility and Sufficiency.* Helsinki: Suomen Kirjallisuuden Seura.

Chlumská, L. (2017). *Překladová čeština a její charakteristiky.* Praha: NLN, Nakladatelství Lidové noviny.

Johansson, M. & Nordrum, L. (2021). Tracing processes in auxiliarization. Time-sufficiency verbs from a Norwegian-Swedish-English contrastive perspective. In A. Čermáková, T. Egan, H. Hasselgård & S. Rør-

vik (Eds.), *Time in Languages, Languages in Time.* Amsterdam: John Benjamins, 67-94.

Kangasniemi, H. (1992). *Modal Expressions in Finnish.* Helsinki: Suomen Kirjallisuuden Seura.

Lindroos, H. A. (1984). *Suomi-Tšekki-Suomi taskusanakirja.* Porvoo: Werner Söderström Osakeyhtiö.

Lingea (2016). *Finsko-český česko-finský šikovný slovník.* Brno: Lingea.

Sinisalo, J. (2011). Weird and proud of it [online]. Transl. D. Hackston, 5. 9. 2011 [cit. 2018-5-10]. Available at: http://www.booksfromfinland.fi/2011/09/weird-and-proud-of-it/.

Tirkkonen-Condit, S. (2002). Translationese - a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target*, 14(2), 207-220.

Tirkkonen-Condit, S. (2004). Unique items? Over-or under-represented in translated language? In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: do they exist?* Amsterdam-Philadelphia: John Benjamins, 177-185.

Corpora

Parallel corpus InterCorp - Finnish and Czech, version 13ud from 18/11/2021. Available at: *InterCorp* corpus ([www.korpus.cz)](www.korpus.cz)

••••

Jonas Freiwald
RWTH Aachen University
jonas.freiwald@ifaar.rwth-aachen.de

Zoe Miljanovic
RWTH Aachen University
zoe.miljanovic@ifaar.rwth-aachen.de

Stella Neumann
RWTH Aachen University
stella.neumann@ifaar.rwth-aachen.de

Arndt Heilmann
RWTH Aachen University
arndt.heilmann@rwth-aachen.de

The effects of entrenchment on translation

This study analyzes how the degree of entrenchment of a grammatical construction affects translation behavior. Entrenchment is understood as a cognitive process of learning a complex structure through repetition leading to its mental representation as a unit, so that "using it is virtually automatic and requires little conscious monitoring" (Langacker 2008, 16). The use of entrenched linguistic patterns is believed to affect the processing speed both during language production and comprehension (Bybee 2010, 34). We assumed that a more frequent construction is also more entrenched in the translator's mind and that different translation solutions are more salient as a result. From a neuro-linguistic perspective, cognitive salience is understood in terms of ease of activation of neural structures and hence likelihood of selection of the corresponding linguistic items. Recent studies have already explored effects of entrenchment as an aspect of cognitive salience on the translation product and process (Halverson 2017); however, further studies are warranted.

Based on these assumptions, we expected that a more frequent construction is translated more quickly than a less frequent construction, resulting in lower reading and typing-related measures of cognitive effort during the translation process. A translation experiment was conducted to test this assumption. Our stimuli consisted of an *of*-NP phrase (cf. Schönthal (2016)), e.g., *the effect of the program*. This partially abstract sequence can realize a

number of linguistic functions, among others possession (*The effect of the program*) and engagement (*The progress of the program*). Schönthal (2016) showed that *of*-NP phrases of possession are the by far most common type of *of*-NP phrases, and we corroborated these results in a corpus study of our own in a subset of the parallel English-German CroCo corpus (Hansen-Schirra et al. 2012). In the 200 instances of our random sample we found 67 cases of possession and 33 cases of engagement, which amounted to 33.5% and 16.5% of *of*-NP phrases overall.

We designed eight short popular scientific texts in a news-ticker message style. Each of the texts included two *of*-NP phrases, one that realizes possession and one that realizes engagement. Each text also came in two versions, where the order of possession and engagement *of*-NP phrases was reversed. Apart from this, the texts were identical in both conditions. Despite the structural equivalence of both of these constructions, we expected less cognitive effort for the possession type of the *of*-NP construction because of the pronounced frequency difference, as it can be expected to be more entrenched and thus dealt with on a more regular basis by professional translators than *of*-NP constructions of engagement.

We asked eleven professional German translators to translate the English original texts into German. We recorded the participants' gaze with the help of a Tobii TX300 eye tracker and the recording software Tobii Studio. We also observed keystrokes, which were logged with Translog-II (Carl 2012). We triangulated the keystroke logging and eye tracking data, which included the measures pause count, translation duration, typing inefficiency, and reading time, and tested the results using linear mixed regression modeling, controlling for, among others, lexically-based effects of salience.

While we did not find statistically significant evidence of a facilitation effect regarding the entrenchment of (partially) abstract grammatical structures (*of*-NPs), we did find salience effects from lexical sources such as cross-linguistic structural priming and words with typical, entrenched translation solutions. Lexical effects of salience on the translation process were shown to be more reliable indicators of facilitation in translation than the salience of more abstract linguistic structures.

References

Book

Bybee, J. (2010). *Language, Usage and Cognition.* Cambridge: Cambridge University Press.

Hansen-Schirra, S., S. Neumann, and E. Steiner (2012). *Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German.* Berlin: De Gruyter.

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction.* Oxford: Oxford University Press.

Schönthal, D. (2016). "On the Multifaceted Nature of English Of-NPs. A Theoretical, Corpus, Contextual and Cognitive Approach". PhD Thesis. Cardiff University.

Article in edited volume

Halverson, S. (2017). "Gravitational Pull in Translation: Testing a Revised Model". In: *Empirical Translation Studies.* Ed. By Gert de Sutter, Marie-Aude Lefer, and Isabelle Delaere. Berlin: De Gruyter, pp. 9-45.

Journal article

Carl, M. (2012). "Translog-II: A Program for Recording User Activity Data for Empirical Reading and Writing Research". In: LREC, pp. 4108-4112.

••••

Chantal Gagnon
Université de Montréal, Canada
chantal.gagnon.4@umontreal.ca

Pier-Pascale Boulanger
Concordia University, Canada
Pierpascale.Boulanger@concordia.ca

A study of "confidence" through financial news translation: a bilingual comparable corpus analysis

Researchers have found that negative information is more compelling and leads to a negativity bias in the mass media (Fuller, 2010; Soroka & McAdams, 2015; Soroka et al., 2018). In sharp contrast with this, business news scholarship has shown that media tend to amplify positive information (Shiller 2008), so as to not spread fear amongst investors (Poole, 2016). Does the overall negativity bias of the media win over the positivity bias of the business news? To answer this question, we surveyed how confidence was reported in the financial press over an eight-year period by French-Canadian and English-Canadian journalists.

News translation is the sub-discipline within translation studies in which this investigation was conducted, not only for the quantitative information that it offers when news reports in two languages are compared, but also for the qualitative insight that is gained through the analysis of divergences and convergences in the journalistic texts produced by two cultural communities. News translation accounts for a complex set of rewriting and editing tasks (Caimotto and Gaspari 2018), which are investigated as instances of intralingual translation, or rewording (Jakobson 1959), a fundamental feature of journalists' work aimed at communicating meaningful versions of world events to large audiences. Using the contrastive approach of translation studies, we searched a 9.2-million word comparable corpus of seven anglophone and francophone Canadian newspapers from 2001 to 2008. The words "confidence" and "confiance" were chosen as gauges of negativity and positivity in the press because confidence is monitored closely by national statistical agencies and central bankers. When it goes up or down, it is a newsworthy event. From the vantage point of news translation studies, we examined how positivity and negativity were communicated by both lin-

guistic communities. This bicultural perspective sheds light on power relations that are represented by the press amongst social groups (i.e. consumers, businesses, investors, central bankers).

This paper will begin by reviewing the scholarship on the negative bias in traditional print media as well as the literature on the positive bias of corporate communication constitutive of business news. It will go on to explain the relation between these findings and news translation research with a focus on critical discourse analysis. After detailing the methodology used to measure negativity and positivity, the paper presents and discusses the results of the corpus study.

The methodology is a hybrid one combining a computer-assisted corpus analysis able to process a high volume of words diachronically and a qualitative approach that provides a critical view of search results. To conduct our research, we used the Canadian Press Corpus in Finance (CAPCOF), composed of Canadian news items covering the 2007-2008 financial crisis and the years that led up to it (2001-2006). According to the Fernandes' typology (2006), CAPCOF is a bilingual comparable corpus. The English corpus (4.8 million words) includes some of the leading dailies in Canada—circulation and reputation wise—, i.e. *The Globe and Mail*, *National Post* and *Toronto Star*. As for the French corpus (4.4 million words), it includes two of the most influential French newspapers in Canada, i.e. *La Presse* and *Le Devoir*. For the sake of representativeness, we have added regional newspapers in each sub-corpora, i.e. the *Montreal Gazette* in English and *Le Droit* in French. The corpus was created with the Factiva database and using keywords often found in the financial pages of major Canadian newspapers between 2001 and 2008 (such as "bank/banque" and "rating agency/agence de notation").

For better workability, including qualitative analysis, the bulk of our research has been focused on a CAPCOF sub-corpus made up of the articles written in 2008. As we will expound on at the conference, it was during this year that the number of occurrences of "confidence/*confiance*" was the highest, frequency peaking at 404 occurrences of "confidence" in English and 442 occurrences of "*confiance*" in French. We analysed manually all the occurrences of "confidence" and "*confiance*" found in this 2008 CAPCOF sub-corpus.

Our partial results show that the words "confidence" and "*confiance*" were used across a complex range of contexts, such as what we label "forecasted positivity" and "forecasted negativity". The presence of forecasting features

in CAPCOF—in the shape of a hypothesis or a prediction by journalists—in both English and French is consistent with findings relating to financial news or financial reporting (e.g., Zuck and Zuck, 1984, Walsh, 2004; Miecznikowski et al., 2012). Our results show that newspapers in both languages were forward-looking in 2008 and viewed the future with relative confidence. However, this positive forecasting is almost twice as frequent in English than in French. This leads us to affirm that the French-language media produced a more balanced ratio of positive/negative occurrences of "*confiance*" than their English-speaking counterparts.

Our paper will contribute to translation studies using corpus-based and corpus-driven methodology in bilingual or multilingual corpora.

References

Caimotto, C., & Gaspari, F. (2018). Corpus-based study of news translation: Challenges and possibilities. *Across Languages and Cultures,* 19(2), 205-220.

Fernandes, L. (2006). Corpora in translation studies: revisiting Baker's typology. *Fragmentos,* 30: 87-95.

Fuller, J. (2010). *What is Happening to News: The Information Explosion and the Crisis in Journalism.* Chicago: University of Chicago Press.

Jakobson, R. (1959). On linguistic aspects of translation. In: *R. Brower (Ed.), On Translation.* Cambridge: Harvard University Press, 232-239.

Poole, R. (2016). Good times, bad times: a keyword analysis of letters to shareholders of two *Fortune 500 banking institutions. International Journal of Business Communication*, 53(1): 55–73.

Miecznikowski, J., Rocci, A., & Zlatkova, G. (2012). L'argumentation dans la presse économique et financière italienne. In: L. Gautier (Ed.), *Les discours de la bourse et de la finance.* Berlin: Frank und Timme, 65-83.

Shiller, R. (2008). *The Subprime Solution.* Princeton: Princeton University Press.

Soroka, S. & McAdams, S. (2015). News, politics, and negativity. *Political Communication,* 32: 1–22.

Soroka, S., Daku, M., Hiaeshutter-Rice, D., Guggenheim, L. & Pasek, J. (2018). Negativity and positivity biases in economic news coverage: Traditional versus social media. *Communication Research,* 45(7): 1078-1098.

Walsh, P. (2004). Investigating prediction in finance and business news articles. In: R. Facchinetti & F. Palmer (Eds), *English Modality in Perspe-*

ctive: Genre Analysis and Contrastive Studies. Frankfurt am Main: Peter Lang, 83–100.

Zuck, J. G., & Zuck, L. V. (1984). Scripts: An example from newspaper texts. *Reading in a Foreign Language,* 21(1): 147-155.

••••

Nataša Gajšt
University of Maribor, Faculty of Economics and Business, Slovenia
natasa.gajst@um.si

A Corpus-Based Study of Syntax-Related Translation Shifts in English-to-Slovenian Translations of Two Types of Legal Texts

Introduction: Translation of legal texts requires the understanding of genre and functional characteristics of legal texts and their translations (Bhatia 1993; Biel 2018; Cao 2007; Gotti 2012; Šarčević 1997; Trosborg 1997) and the knowledge of syntactic structures in source and target languages (Halliday 2001; Hiltunen 2012). Complex syntactic structures may present a translation challenge from the linguistic viewpoint and from the viewpoint of maintaining these texts' meta-functions (Matthiessen 2001). According to Catford's (1965) theory of translation, structural differences between pairs of languages result in structure and unit shifts in the syntactic configuration of the target text compared with the source text. Typical structure shifts are the changes in the order of clause/sentence elements in the target language (TL) clause/sentence structure compared to source language (SL) clause/sentence structure. An example of unit shift at clause/sentence levels is a change from a simple sentence in a SL to a complex sentence in a TL. Frequently, unit shifts cause structure shifts in TL sentences.

Objectives: The first aim of our research was to establish different types of structure and unit shifts at sentence/clause levels in translations of two types of legal texts from English to Slovenian, i.e., international/bilateral treaties and the EU Court of Justice judgements. Secondly, we aimed to determine whether the scope of the established shifts in translations can be ascribed to the genre and functional differences of these two types of legal texts or not.

Method: Our study of structure and unit shifts integrated the principles of translation-oriented contrastive analysis and corpus linguistics (Catford 1965; James 1980; Borja, Izquierdo, & Montalt 2009; Gorjanc 2005; McEnery & Xiao 2007). We built a small-scale specialized one-directional parallel English-Slovenian corpus comprising selected international/bilateral treaties signed by the Republic of Slovenia and selected EU Court of Justice

judgements. The corpus was based on an equal number of sentences, i.e., 200 sentences per each sub-corpus. The criteria for including texts into corpora were based on textual and genre characteristics of these types of texts (Mikhailov & Cooper 2016). We developed an annotation schema based on the theory of syntactic structures in English and Slovenian and on the series of detailed probe tests and questions for determining sentence elements, which included functional annotation, too (Biber et al. 1999; Burton-Roberts; 2011; Toporišič 1982, 1991; Berry 2019; Eggins 2004; Fontaine 2012; Fawcett 2000; Halliday & Matthiessen 2004; Morley 2000, 2004; Thompson 2014). The author of this research manually annotated syntactic structures in sentences and clauses in an electronic spreadsheet program. We then performed a quantitative analysis and qualitative analysis (Hasko 2013) of syntactic structures and syntax-related translation shifts for each type of legal text.

Results: Our empirical analysis confirmed that the differences between English and Slovenian linguistic systems lead to structure and unit shifts at sentence and clause levels. We found that the scope of shifts was about twice as large at the clause level than at the sentence level, which supports Catford's (1965) premise that more shifts occur at lower ranks than at higher ones. Regarding structure shifts at sentence level, we established that they occurred mainly due to the shifted position of the adverbial phrase, the changes of clause elements order due to passive-to-active sentence structure shift, or due to the verb phrase being embedded with other clause elements. For instance, English structure DO+P+A+S typically changed to Slovenian structure DO+P<S>+A where P was embedded by S. A number of structure shifts at both clause and structure levels were due to unit shifts. Concerning unit shifts, we found they were prevailingly the result of complex, extensively post-modified English noun phrases functioning as objects. These structures typically evolved into subordinate clauses in Slovenian. For example, English simple sentence structure S+P+DO with head in DO postmodified by a complex ed-clause changed to Slovenian complex sentence structure S+P+DO (the main clause) + subordinate clause where Slovenian subordinate clause evolved from the English ed-clause as postmodifier to head in DO. The analysis of the scope of different shifts per two types of legal texts showed that the same types of structure shifts and unit shifts occurred in translations of both types of texts and also in a somewhat similar scope. The difference in the scope of the analysed translation shifts at sentence level between the two sub-corpora was only 5 percent.

Conclusion and recommendations: We can conclude that translation shifts determined in our study are due to differences between English and Slovenian language systems rather than the differences in genre and functional characteristics of international/bilateral treaties and the EU Court of Justice judgements. Future contrastive and/or comparative analyses of translation shifts can be done for translations of texts from different language pairs and/or different professional fields.

Keywords: legal translation, complex syntactic structures, translation shifts, contrastive analysis, international treaties, bilateral treaties, EU Court of Justice judgements, English, Slovenian

References

Berry, M. (2019). The clause: an overview of the lexicogrammar. In G. Thompson, W. L. Bowcher, L. Fontaine & D. Schönthal (Eds.), *The Cambridge Handbook of Systemic Functional Linguistics*. Cambridge: Cambridge University Press, 92-117.

Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. Harlow: Longman.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow (Essex): Longman.

Biel, Ł. (2018). Genre analysis and translation. In K. Malmkjær *(Ed.), The Routledge Handbook of Translation Studies and Linguistics*. London: Routledge, 151-164.

Borja, A., Izquierdo, I. G., Montalt, V. (2009). Research methodology in specialized genres for translation purposes. *The Interpreter and Translator Trainer*, 3 (1), 57-77.

Burton-Roberts, N. (2011). *Analysing Sentences: An Introduction to English Syntax*. Harlow: Pearson Education Limited.

Cao, D. (2007). *Translating Law, Topics in Translation*. Clevedon: Multilingual Matters.

Catford, J. C. (1965). *A Linguistic Theory of Translation*. Oxford: Oxford University Press.

Eggins, S. (2004). *Introduction to Systemic Functional Linguistics*. London: Continuum International Publishing Group.

Fawcett, R. (2000). *A Theory of Syntax for Systemic Functional Linguistics*. Amsterdam: John Benjamins.

Fontaine, L. (2012). *Analysing English Grammar: A Systemic Functional Introduction*. Cambridge: Cambridge University Press.

Gorjanc, V. (2005). *Uvod v korpusno jezikoslovje*. Domžale: Izolit.

Gotti, M. (2012). Text and genre. In L. M. Solan & P. M. Tiersma (Eds.), *The Oxford Handbook of Langauge and Law*. Oxford: Oxford University Press, 52-66.

Halliday, M.A.K, Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar*. London: Arnold.

Halliday, M. A. K. (2001). Towards a theory of good translation. In E. Steiner & C. Yallop (Eds.), *Exploring Translation and Multilingual Text Production: Beyond Content*. Berlin, New York: De Gruyter Mouton, 13-18.

Hasko, V. (2013). Qualitative corpus analysis. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics,* Malden, MA: Blackwell Publishing Ltd., 4758-4764.

Hiltunen, R. (2012). The grammar and structure of legal texts. In L. M. Solan & P. M. Tiersma (Eds.), *The Oxford Handbook of Langauge and Law*. Oxford: Oxford University Press, 39-51.

James, C. (1980). *Contrastive Analysis*. London: Longman.

Matthiessen, C. M. I. M. (2001). The environments of translation. In E. Steiner & C. Yallop (Eds.), *Exploring Translation and Multilingual Text Production: Beyond Content*. Berlin: De Gruyter Mouton, 41-124.

McEnery, T., Xiao, R. (2007). Parallel and comparable corpora: what is happening? In G. Anderman & M. Rogers (Eds.), *Incorporating Corpora*. Clevedon: Multilingual Matters, 18-31.

Mikhailov, M, Cooper, R. (2016). *Corpus Linguistics for Translation and Contrastive Studies: A Guide for Research*. Oxon: Routledge.

Morley, G. D. (2000). *Syntax in Functional Grammar: An Introduction to Lexicogrammar in Systemic Linguistics*. London: Continuum.

Morley, G. D. (2004). *Explorations in Functional Syntax: A New Framework for Lexicogrammatical Analysis*. London: Equinox Publishing Ltd.

Šarčević, S. (1997). *New Approach to Legal Translation*. Haag: Kluwer Law International BV.

Thompson, G. (2014). *Introducing Functional Grammar*. Oxon: Routledge.

Toporišič, J. (1982). *Nova slovenska skladnja*. Ljubljana: Državna založba Slovenije.

Toporišič, J. (1991). *Slovenska slovnica*. Maribor: Založba Obzorja.

Trosborg, A. (1997). Text typology: register, genre and text type. In A. Trosborg (Ed.), *Text Typology and Translation*. Amsterdam: John Benjamins, 3-24.

••••

Volker Gast
Friedrich-Schiller Universität Jena, Germany
volker.gast@uni-jena.de

Vahram Atayan
Institute for Translation and Interpreting, Heidelberg University
atayan@uni-heidelberg.de

Visualizing multi-dimensional data in translation studies and contrastive linguistics

The availability of large amounts of multilingual corpus data has opened up new possibilities for empirical approaches to translation studies and contrastive linguistics. For example, the corpus resources available today have enabled multivariate studies (e.g. regression modeling, cf. Gries & Wulff 2011). We can empirically investigate the conditions under which specific expressions are translated in a specific way, we can determine the contexts in which comparable linguistic expressions from different languages are used, etc. (e.g. Gast 2015, Atayan 2021). These possibilities come with new challenges, however. From the perspective of the researcher, it is sometimes hard to filter out more relevant from less relevant information, to understand the interplay of the variables used in a study, and to interpret the results in a theoretically meaningful way, beyond statements of the type „variable x has a significant effect on variable y".

While inferential statistics is indispensable in the comparative analysis of corpora, it is important for researchers and readers to also gain an intuitive understanding of the data. Visualizations are an important tool for this purpose. In this contribution we address the following questions relating to the visualization of corpus data in translation studies and contrastive linguistics:

- What visualization techniques are there, for what type of data, and what type of „problem"?
- What strengths and weaknesses do the various types of visualization techniques have, relative to a given objective?

We will briefly discuss some standard methods of visualization like the following:

- visualization of relative over- and underrepresentation of variable combinations (e.g. mosaic and association plots based on Chi-squared tests; feature combination frequency grids [FCF-grids]);
- visualization of correlations (e.g. ordination techniques such as Principal Component Analysis [PCA], Correspondence Analysis [CA], Multiple Correspondence Analysis [MCA]);
- visualization of statistical models (e.g. regression models with random effects, conditional infererence trees, random forests).

One focus of our presentation will be on a family of visualization techniques that has not so far been widely used, i.e., techniques making use of graphs (in a mathematical sense) — sets of nodes and edges connecting the nodes. While graphs do not make use of (Euclidean) space to represent degrees of similarity or distance — like ordination techniques such as PCA, CA and MCA — they offer other possibilities, e.g. insofar as edges can represent properties of relations holding between nodes in terms of color, thickness and length. Specifically, we will present two types of graphs for the representation of translated data: (i) *context-conditional translation graphs*, which illustrate what type of translation is found under what conditions, with what frequencies, in a specific corpus; and (ii) *context-conditional correlation graphs*, which show degrees of difference between linguistic expressions from different languages in terms of the type of context in which they occur .

In our presentation we will discuss the strengths and weaknesses of the different techniques. For example, among the ordination techniques for categorical variables, CA is particularly useful if one wants to make inferences from expressions to contexts, whereas MCA is better suited for inferences from contexts to expressions. Similarly, context-conditional translation graphs visualize the data by expression, whereas context-conditional correlation graphs do so by context. We will also address the question of visualizing neural networks and their internal structure.

References

Atayan, Vahram (2021). Marker der Nachzeitigkeit im deutsch-italienischen Vergleich. Eine korpusbasierte Studie zu *gleich, sofort, immediatamente und subito. In Katelhön, Peggy and A. Muco (eds.), Deutsche Sprache im Kontrast. Sonderheft von Linguistik Online*, 11-43, https://bop.unibe.ch/linguistik-online/article/view/8234/11230.

Gast, Volker (2015). On the use of translation corpora in contrastive linguistics. A case study of impersonalization in English and German. *Languages in Contrast* 15.1: 4-33.

Gast, Volker, Vahram Atayan, Julien Biege, Bettina Fetzer, Sophie Hettrich and Anne Weber (2019). Unmittelbare Nachzeitigkeit im Deutschen und Französischen: Eine Studie auf Grundlage des OpenSubtitles-Korpus. In Lavric, Eva et al. (eds.): *Comparatio delectat III. Akten der VIII. Internationalen Arbeitstagung zum romanisch-deutschen und innerromanischen Sprachvergleich*, 223–250. Wien, Lang.

Gries, Stefan Th. and Stefanie Wulff (2011). Regression analysis in translation studies. In Oakes, Michael P. and Meng Ji (eds.), Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research, 35-53. Amsterdam: Benjamins.

Volker Gast
Friedrich-Schiller Universität Jena, Germany
volker.gast@uni-jena.de

Robert Borges
Polish Academy of Sciences
robert.d.borges@gmail.com

A triangular translation corpus, and a case study of nominal compounds in English and German

Large amounts of data and new technologies and infrastructures have opened up new possibilities for the study of translation and interpreting. In this talk we present a multimodal corpus that contains video files of speeches made in the European Parliament, with transcriptions of the speeches themselves as well as 21 simultaneous interpretations (similar to EPTIC, cf. Bernardini et al. 2016). The corpus is stored in EAF-files for ELAN (Sloetjes & Wittenburg 2008), the speeches themselves are provided in MP4-files. The corpus also contains written translations of the speeches. We call this type of corpus „triangular" because for each speech, there are three types of text: the original, the (written) translations and the (spoken) interpretations.

A trianguar translation corpus allows us to study a number of interesting issues in a new light. For example, translations and interpretations of the same speech into different languages can be compared, with an essentially typological research interest. In this talk we will use the corpus for another purpose: we will compare translations and interpretations of the same texts, using English and German data. The main intention is to gain insights into bilingual cognition under different input and output conditions (see for instance Defrancq and Plevoets 2018 for a study of bilingual cognition in interpreting).

The domain of inquiry is nominal compounding, an area in which English and German show considerable differences (cf. Berg 2017, Zeldes 2018; see König & Gast 2018: Ch. 8 for a comparative overview). In English, compounds are mostly defined on the basis of phonological and semantic criteria (see for instance Plag 2003), and they constitute a relatively „fuzzy" category, in the sense that they cannot always straightforwardly be distinguished from syntactic phrases. Moreover, nominal compounds in English tend to be

binary, i.e., they mostly consist of two components. In German, compounding has morphological reflexes (Fuhrhopp 1998) and complex compounds (with more than two components) are not uncommon. There is thus reason to assume that compounds may have different mental representations in English and German, insofar as in German they have a „more lexical" status than in English. This, in turn, suggests that the distribution of labour between lookup and parsing operations may differ.

We will compare translations and interpretations of both English speeches translated into German, and German speeches translated into English, on the basis of a random sample of 150 speeches for each direction of translation. The expectation is that specific types of „mismatches" between the source and target structures will lead to different types of processing difficulties. We treat the written translations, which were made without time pressure, as „gold standards", and identify deviations from those standards (e.g. omissions and non-idionatic translations) in the simultaneous interpretations (under time pressure). We assume that the amount of deviation between written translations and interpretations reflects processing difficulties. For the quantitative analysis we will use multilevel regression modeling.

Some of the discrepancies observed between translations and interpretations are illustrated in Table 1, from a speech given by Richard Howitt on 14 January, 2009, in the European Parliament. While the rare expression *the going down of the sun* is left untranslated in the interpretation (the written translation has *Sonnenuntergang*), and *massacre* and *genocide* are interpreted using the cognates *Massaker* and *Genozid* (vs. the compounds *Blut-bad* ‚blood-bath' and *Völker-mord* ‚people-murder' in the written translation), both types of translations have *Sicherheitsrat* (‚Security Council') and *Rassenhass* (‚racial hatred'). We assume that translation difficulties as reflected in mistranslations and longer reaction times (lags) in interpreting are a function of parallelisms or mismatches in structure, frequency of occurrence and salience in a given context. From the perspective of bilingual cognition we assume that lookup-lookup pairs of the type ⟨*Security Council*, *Sicherheitsrat*⟩ come with low processing cost whereas parsing-lookup mismatches of the type ⟨*the going down of the sun*, *Sonnenuntergang*⟩ demand higher processing efforts.

English original	German written translation	German simultaneous interpretation
the going down of the sun	Sonnenuntergang	[not translated]
human cost of war	Menschenleben, die ein Krieg kostet	die Kosten des Krieges
massacre	Blutbad	Massaker
United Nations Security Council	Sicherheitsrat	Sicherheitsrat
racial hatred	Rassenhass	Rassenhass
networking site	Netzwerkseite	[not translated]
genocide	Völkermord	Genozid

Table 1: Compounds in a speech given by Richard Howitt on 14 January, 2009

References

Berg. T. (2017). Compounding in German and English. *Languages in Contrast* 17 (2): 43-68.

Bernardini, S., A. Ferraresi and M. Miličević (2016). From EPIC to EPTIC – Exploring simplification in interpreting and translation from an inter-modal perspective. *Target* 28: 61-86.

Defrancq, Bart and Koen Plevoets (2018). Over-uh-Load, Filled Pauses in Compounds as a Signal of Cognitive Load. In Russo et al. (eds.), *Making Way in Corpus-Based Interpreting Studies*, 43–64. Singapore: Springer.

Fuhrhopp, N. (1998). *Grenzfälle morphologischer Einheiten*. Tübingen: Stauffenburg.

König, E. and V. Gast (2018). *Understanding English-German Contrasts. 4th edition. Berlin: Erich-Schmidt Verlag.Plag, I. (2003). Word Formation in English.* Canbridge: Cambridge University Press.

Zeldes, A. (2018). Compounds and productivity in advanced L2 German writing. In Tyler, A.E., L. Ortega, M. Uno and H.I. Park (eds.), *Usage-Inspired L2 Instruction: Researched Pedagogy*, 237-265. Amsterdam: Benjamins.

Sloetjes, H. and P. Wittenburg (2008). Annotation by category - ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.

••••

Patrizia Giampieri
University of Camerino, Italy
patrizia.giampieri@unicam.it

The drafting style of Employment Agreements and Employment Offer Letters – a Corpus-based approach

Introduction

Legal language is considered pedantic and archaic (Bhatia 1997; Williams 2004 and 2011). Scholars claim that the "language of the law" (Mellinkoff 2004) is intricate for reasons of inclusiveness and preciseness (Coulthard and Johnson 2007: 40). Others, instead, argue that lawyers and judges do not have any particular communication intent; they use a complicate jargon only to remark their "legal fraternity" (Tiersma 1999: 52). Many legal documents and deeds are still drawn up in the most intricate form, which is referred to as *legalese.* Even international contracts addressed to non-native speakers of English (such as employment agreements) tend to be drafted in an archaic style. In this respect, there seems to be a significant difference between employment agreements and employment offer letters. This paper is aimed at bringing to the surface the nuances, if any, between these two types of contracts. Therefore, it explores the language of employment contracts and employment offer letters by unveiling differences in their legal language and style. In particular, it analyses whether these documents present elements of *legalese* and/or plain English. In order to do so, two corpora (one of employment agreements, the other of employment offer letters) are built, and the two contracts are explored and analysed in light of corpus evidence. Given its specificity and technicalities, in fact, legal language has been increasingly studied and analysed by using and consulting corpora (Vigier Moreno and Sánchez Ramos 2017; Biel 2018; Giampieri 2018; Vigier Moreno 2019; Giampieri and Milani 2021). Corpora help find and notice word usages in context, collocations and formulaic expressions, which abound in legal texts.

Methodology

In order to investigate the legal language and style of employment contracts and employment offer letters, two specific corpora are built: a corpus of em-

ployment agreements and a corpus of employment offer letters. The corpora are created manually in order to make sure that the documents retrieved are written in English as a first language. To guarantee this important aspect, advanced search settings are applied.

Results

The analysis shows a recurrent and systematic use of *legalese* in employment agreements. Employment offer letters, instead, seem to be hallmarked by *legalese* to a lesser extent. There are yet long sentences, complex prepositions, passive forms and nominalisation, but *legalese* elements are less pervasive; texts are smoother and, hence, more understandable.

Conclusions

This paper contributed to highlighting the fact that employment agreements are still drafted in a formal and archaic style, despite being addressed to non-native speakers of English. In employment letter offers, instead, alternative (or plainer) linguistic solutions are proposed and texts are organised in bullet points.

Differently from what could be expected, the findings of this paper highlight that employment agreements and employment offer letters do not differ in terms of the duties or positions offered to the recipient, as they can be both addressed to executives or directors. Hence, the reason(s) why companies decide to resort to one type of document instead of the other might be further investigated. Also, analysing the drafting style of contracts written in English as a second language could be another avenue to explore.
The implications of this study in translation training and practice are manifold, as this paper provides evidence that corpus consultation improves the quality of translation and helps produce native-like target texts, care being taken to consider the (dis)similarities between the source and target legal systems of reference..

References

Bhatia, V. K. (1997). Translating Legal Genres. In A. TROSBORG (ed), *Text Typology and Translation*. Amsterdam: John Benjamins, 203-214.

Biel, Ł. (2018). Corpora in Institutional Legal Translation: Small Steps and the Big Picture. In F. Prieto Ramos (Ed), *Institutional Translation for Inter-*

national Governance: Enhancing Quality in Multilingual Legal Communication. London: Bloomsbury, 25-36.

Coulthard, M. and Jhonson, A. (2007). *An Introduction to Forensic Linguistics. Language in Evidence.* London/New York: Routledge.

Giampieri, P. (2018). Online parallel and comparable corpora for legal translations. *Altre Modernità*, 20, 237-252.

Giampieri, P. And Milani, G. (2021). Legal Corpora: A Trial Lesson With Translators And Lawyers, *International Journal Of Language Studies*, 15(1), 47-66.

Mellinkoff, D. (2004). *The language of the law.* Eugene: Resource Publications.

Tiersma, P. (1999). *Legal Language.* Chicago, Il: The University Of Chicago Press.

Vigier Moreno, F. J. (2019). Corpus-Assisted Translation Of Specialised Texts Into The L2. From The Classroom To Professional Practice. *Trans-Kom,* 12 (1), 90-106.

Vigier Moreno, F. J. and Sánchez Ramos, M. (2017). Using Parallel Corpora to Study the Translation of Legal System-Bound Terms: The Case of Names of English and Spanish Courts. In: R. Mitkov (ed), *Computational and Corpus-Based Phraseology EUROPHRAS 2017, Lecture Notes in Computer Science*, 10596. Cham: Springer, 206-273.

Williams, C. (2004). Legal English And Plain Language: An Introduction. *Esp Across Culture*, 1, 111-124.

Williams, C. (2011). Legal English And Plain Language: An Update. *Esp Across Cultures*, 8, 139-151.

••••

Kiara Giancola
University of Mons
kiara.giancola@student.umons.ac.be

Towards the Development of an Error Typology in Sight Translation

Sight translation can be defined as "the oral rendition of a written document in the target language" (Mellinger, 2017, p. 312). This practice has gained momentum as a valuable pedagogical tool in the realm of simultaneous and consecutive interpreting (Li, 2014). However, based on the European Master's in Translation (European Commission, n.d.), only seldom has this type of translation been incorporated into translation training programs in Europe. Yet, according to Dragsted & Hansen's study (2009), which compared written and sight translation, sight translation should be considered a relevant teaching tool within translation training, since it saves more time than written translation, without compromising the quality of the output. Nevertheless, research in this area remains underexplored.

This present paper is an attempt to investigate sight translation in the realm of translation training, with a particular focus on translation mistakes. Indeed, error analysis holds a central place in translation studies (Soltani et al., 2020), but only few researchers have worked on sight translation errors (Li, 2014). Accordingly, this research aims to put forward a thoroughly explained error typology that perfectly fits sight translation so as to examine the trends specifically tied to this practice.

The analysis is based upon a corpus called "Transcription of 20 sight translations (EN>FR) by translation students coded with voiced pauses and silent pauses" (Meyers, 2022). In order to build this corpus, twenty translation students were asked to sight translate a 433-word article from the New Scientist called "Looking into the Voids Could Help Explain Dark Energy" (Ananthaswamy, 2015). During the exercise, the participants were recorded while sight translating. Their output was then compiled and transcribed. As part of this present research, transcriptions were analyzed to spot mistakes and classify them in accordance with an error typology.

The typology is inspired by different existing typologies, namely the Multidimensional Quality Metrics (MQM) (Lommel et al., 2015) and Falbo's ty-

pology (1998). The relevant categories of these two were reworked and used as a foundation. However, no matter how developed these typologies are, some mistakes did not match any existing categories. To tackle this issue, new categories were created based on the analyzed output. These focus on a further detailed division of mistranslation errors that can be encountered in a text: opposite meaning (i.e., the target content conveys the opposite meaning of the source content), mistranslation (i.e., the target content slightly differs from the source content), non-sense (i.e., the target content makes no sense) and lexical mistranslation (i.e., use of the wrong acceptation of a substantive based on the context of the source content). Moreover, the analysis led to the creation of four adjacent categories based on Shreve et al.'s typology (2011): unfilled pause, filled pause, repetition, and repair. Unfilled pauses appear whenever there are "interruptions in speech that do not contain any phonetic content" (Shreve et al., 2011, p. 99); filled pauses are "interruptions in speech that are filled with utterances such as [euh]" (Shreve et al., 2011, p. 99); repetitions are detected whenever "the speaker repeats part of the utterance" (Fox Tree, 1995, p. 710); and repairs are considered when "an utterance is repeated, but with an alteration to one or more constituents (for instance by replacement, omission, or insertion of words), and the other constituents generally preserved" (Shreve et al., 2011, p. 99; definition and terminology inspired by Shriberg, 1994).These aspects do not directly act as error categories *per se*, but rather as a gauge to determine whether speech disfluencies have any influence on the total number of errors made by the students. All in all, the combination of existing and new error categories, as well as the four adjacent aspects, contributed to the completion of an error typology tailor-made for sight translation.

All aspects considered, these results may prove to be useful in the field of both translation research and translation teaching. It could provide researchers with a better-fitted typology for the analysis of sight translation corpora and help contribute to moving sight translation research forward. This typology could also be relevant to further investigate the promising role of sight translation in translation courses.

References

Ananthaswamy, A. (2015, April 8). Looking into the voids could help explain dark energy. *New Scientist*, https://www.newscientist.com/article/mg22630164-600-lookinginto-the-voids-could-help-explain-dark-energy/

Dragsted, B., & Hansen, I. (2009). Exploring Translation and Interpreting Hybrids. The Case of Sight Translation. *Meta*, 54 (3), 588–604. https://doi.org/10.7202/038317ar

European Commission. *List of EMT members 2019-2024*. (n.d.). Retrieved April 20, 2021, from https://ec.europa.eu/info/resources-partners/european-masters-translation-emt/list-emt-members-2019-2024_en

Falbo, C. (1998). Analyse des erreurs en interprétation simultanée. In EUT Edizioni Università di Trieste (ed.), *The Interpreters' Newsletter*, 8/1998, pp. 107-120.

Fox Tree, J. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34, 709–738.

Li, X. (2014). Sight translation as a topic in interpreting research: Progress, problems, and prospects. *Across Languages and Cultures*, 15(1), 67-89. Retrieved Sep 8, 2021, from https://akjournals.com/view/journals/084/15/1/article-p67.xml

Lommel, A., et al. (ed.). (2015). *Multidimensional Quality Metrics (MQM) Definition*. German Research Center for Artificial Intelligence and QTLaunchPad, Retrieved August 9, 2021, from http://www.qt21.eu/mqm-definition/definition-2015-12-30.html

Mellinger, C. D. (2017). Equal access to the courts in translation: A corpus-driven study on translation shifts in waivers of counsel. *Perspectives*, 25(2), 308–322. https://doi.org/10.1080/0907676X.2016.1248985.

Meyers, C. (2022). *Transcription of 20 sight translations (EN>FR) by translation students coded with voiced pauses and silent pauses* [Data set]. Social Sciences and Digital Humanities Archive – SODHA. https://doi.org/10.34934/DVN/KHQE0P

Shreve, G. M., Lacruz, I. and Angelone, E. (2011). Sight Translation and Speech disfluency: Performance Analysis as a Window to Cognitive Translation Processes. In B. Englund Dimitrova, C. Alvstad, A. Hild and E. Tiselius (eds.), *Methods and Strategies of Process Research: Integrative Ap-*

proaches in Translation Studies, John Benjamins Publishing Company, 93-121.

Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies* [Unpublished doctoral dissertation]. University of California.

Soltani, F., Nemati, A., & Yamini, M. (2020). An analysis of translation errors in 5 literary genres based on American Translation Association (ATA) framework. *Cogent Arts & Humanities, 7*(1). https://doi.org/10.1080/233 11983.2020.1799732

••••

Tora Hedin
Stockholm University, Sweden
tora.hedin@slav.su.se

Irene Elmerot
Stockholm University, Sweden
irene.elmerot@slav.su.se

The translation of Czech particles in three registers: a corpus-based analysis

The present study will analyze Czech particles and their Swedish parallels in different registers. The study is carried out within a larger project where we have wanted to distinguish between Czech as source language (SL) and Czech as target language (TL) in the analysis.

Particles and their functions are a challenge for researchers. Particles have weak semantics, serve a pragmatic function and are often polysemic, which makes them hard or sometimes impossible to translate literally. At the same time, they can be crucial to our interpretation of an utterance to such a degree that the meaning would be completely altered if a particle is removed or changed.

The use of Czech particles will be compared in three registers: fiction, spoken political texts and legislative texts, and their Swedish translation parallels will be analyzed. The empirical data have been collected from the Inter-Corp14 Czech[1] and InterCorp14 Swedish: texts of fiction (Core), transcripts from proceedings from the European parliament (EUROPARL; Koehn, 2005) and ACQUIS, legislative EU texts. Three particles will be presented here: *právě* ('just', 'right', 'exactly'), *bohužel* ('unfortunately') and *asi* ('about', 'or so'; 'apparently'). In two of the subcorpora, Czech and Swedish are SL, whereas in the third (ACQUIS) both Czech and Swedish are translated. In EU, all official languages enjoy equal status (European Parliament 2021), and the ACQUIS translations follow certain rules that justify their inclusion.[2]

---

1   korpus.cz: https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze14

2   This circumstance will be discussed further in the study itself.

Contrastive analysis using parallel corpora is a growing research field. Studies of Czech and Swedish particles, respectively, have been published earlier, but not between the two languages. In a Swedish context, Aijmer (i.e. 2007) has conducted a number of studies of equivalents between Swedish and English. In a Czech context, Šebestová and Malá, using InterCorp, found that the postfix *-pak* often has interrogative sentences in translations to English. Martinková and Janebová (2017) has showed difference between three registers for the translation of the particle *prý* into English. Worth mentioning is also Šindlerová and Štěpánková (2021).

There is so far no unified categorization of Czech particles, but criteria such as function and form (Nekula 2017) have been used. In Swedish grammar particles have been classified as adverbs, interjections, subjunctions, etc. (Teleman et al. 1999:207). For the present analysis, we use Čermák (2008:64–65): modal/epistemic particles, emotional particles, evaluative particles and intensifying particles.

We will present some results and 2 examples from the corpora to illustrate our arguments. Quantitative analysis shows that all three particles tend to be much more overall dominant, seen as i.p.f, in fiction (Core) when the source language is Swedish, and more dominant in EUROPARL when the source language is Czech.

Tentative results show that the Czech particles are often represented by a large number of Swedish equivalents in the corpora. However, the Czech particles *právě*, *bohužel* and *asi* differ in their function, the number of Swedish equivalents and translation strategies between the registers.

The emotional particle *bohužel* is generally translated to or from the equivalent *tyvärr* ('unfortunately'). The number of equivalents is lower for this particle than for the two others. Zero equivalence is, however, more frequent in EUROPARL (CZ-SV) than in the other corpora

The particle *právě* has a number of equivalents in the Swedish translations in FICTION, but one of them, *just*, is fairly dominant. In EUROPARL, *právě* is again often omitted, changing the meaning of the utterance, cf. (2):

  (1) *Právě* minimální standardy by měly jejich bezpečnost zajistit.
  'Minimistandarder bör garantera säkerheten för konsumenterna.'
  [Minimum standards should ensure consumer safety.]
  (EUROPARL)

There is a large number of translation equivalents in FICTION of *asi*, such as *nog, ungefär* ('about'), *antagligen* ('probably'), *omkring* ('around') or verbal constructions, such as *jag antar* ('I presume'). EUROPARL again shows more zero equivalence. We also find a different semantics in the two other corpora compared to the translated works of fiction; illustrated by *asi* being translated as *uppenbarligen* ('obviously/apparently'), cf. (3):

(2)
Asi nám nezbývá, než prodloužení odhlasovat, protože by jinak řada přípravků musela z trhu zmizet.
'Vi hade *uppenbarligen* inget annat val (…)
[*Apparently* we had no choice (…)]
(EUROPARL)

The preliminary results show that there is a vast variation in equivalents in the translation of Czech particles in Fiction. In Acquis, the Czech particles seem to be absent and in Europarl, they are often omitted in translations, and the number of equivalents is less than in Fiction. Our study confirms that there is indeed a difference between registers regarding the function of particles and the translational strategies used.

A possible explanation can be the universal tendency to simplify in TL compared to SL (see Cvrček and Chlumská, 2015). In the case of Europarl, this may partly be due to the Interinstitutional style guide of the EU (Publications Office 2021). Europarl consisting of transcripts from oral language is also a factor.

References

Aijmer, K. (2007). The meaning and functions of the Swedish discourse marker *alltså – Evidence from translation corpora. Catalan Journal of Linguistics* 6, 2007 31-59.
Aijmer, K. & Simon-Vandenbergen, A.-M. (2004). A model and a methodology for the study of pragmatic markers: the semantic field of expectation. *Journal of Pragmatics*, 36(10), 1781–1805.
Cvrček, V. et al. (2010). *Mluvnice současné češtiny 2*. Karolinum.

Cvrček, V., Chlumská, L. (2015). Simplification in translated Czech: a new approach to type-token ratio. *Russian Linguistic* **39,** 309–325. https://doi.org/10.1007/s11185-015-9151-8

European Parliament. (2021). *Rules of Procedure of the European Parliament.* https://www.europarl.europa.eu/doceo/document/lastrules/RULE-167_EN.html

Martinková, M. & Janebová, M. (2017). What English Translation Equivalents Can Reveal about the Czech "Modal" Particle prý: A Cross-Register Study. In: Aijmer, K. & Lewis, D. (eds.) *Contrastive Analysis of Discourse-pragmatic Aspects of Linguistic Genres* Yearbook of Corpus Linguistics and Pragmatics. [Online]. Cham: Springer International Publishing.

Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In: Hutchins, J. (ed.) The Tenth Machine Translation Summit Conference Proceedings. [Online]. International Association for Machine Translation, pp. 79-86. Available at: https://www.research.ed.ac.uk/en/publications/europarl-a-parallel-corpus-for-statistical-machine-translation;

Nekula, M. (n.d.) Částice. *CzechEncy - Nový encyklopedický slovník češtiny.* Online. Available at: https://www.czechency.org/slovnik/ČÁSTICE [Accessed: 14 February 2022].

Publications Office. (2021). *Interinstitutional style guide.* [Online]. Luxembourg: Publications Office. Available at: https://data.europa.eu/doi/10.2830/611377 [Accessed: 14 February 2022].

Šebestová, D. & Malá, M. (2016). Anglické překladové protějšky českých vět s částicemi copak a jestlipak. *Časopis pro moderní filologii*, 98(2), 228–237.

Šindlerová, J. & Štěpánková, B. (2021). Between adverbs and particles: A corpus study of selected intensifiers. *Journal of Linguistics/Jazykovedný časopis*, 72(2), 444–453.

Teleman, U. et al. (1999). *Svenska akademiens grammatik.* [Online]. Stockholm: Svenska Akademien / Norstedts. Available at: https://svenska.se/SAG.pdf [Accessed: 14 February 2022].

••••

Attila Imre
Sapientia Hungarian University of Transylvania, Cluj-Napoca, Romania
attilaimre@ms.sapientia.ro

Machine translation support for frequent acronyms of American TV series focusing on Hungarian as the target language

The aim of our research is to test whether present-day MT can offer good results in the case of acronyms for the English–Hungarian language pair. It is known that the frequency of acronyms is underrated (Izura & Playfoot, 2012: 869), although – interestingly – they became popular in the 1940s (Yeates, 1999: 118), and their use reached "epidemic proportions" (Jamieson, 1968: 473) in all fields in about two decades. Today we witness a "compulsive use of acronyms" (Begg, 2017: 561) not only in scientific papers but on social media as well. While we presume that MT cannot decipher acronyms, it can still spot them and preserve them, which is a rather frequent translation strategy for human translation as well. Yet, we tend to believe that a few popular English acronyms might have their Hungarian renditions included in the MT database, some of which are borrowed (*CV, DJ, OK* and *TV*), while others have established equivalents, which are not necessarily acronyms (*CDC* or *FBI*).

Algorithms developed to collect acronyms and their extended versions only focus on monolingual corpora in specialized journals (especially legal or medical) or the Internet (Barnett & Doubleday, 2021; Cannon, 1989; Zahariev, 2004), so we could not find any research including both algorithms to track them and their translation. Although "matching on anything less than 3 characters is very error prone" (Taghva & Gilbreth, 1999: 192), we have created an algorithm to identify acronyms based on a string of at least two uppercase letters with or without periods between them. *Acronym* is used as an umbrella term for abbreviations, acronyms, initialisms or alphabetisms, as neither linguistic approaches nor algorithm-driven research could define acronyms unequivocally, not to mention that they constitute a real challenge for translators and subtitlers alike.

The corpus is based on 586 English subtitles of five popular American TV series (*The West Wing, 24, House of Cards, Blindspot* and *Designated Sur-*

vivor), resulting in almost 1,500 unique values and almost 13,000 acronyms, without counting the 'false alarms' (full uppercase letter words to highlight important pieces of information in subtitles, such as names, labels, tags, etc.), which were excluded after the preliminary results. This means that – on average – slightly more than 22 acronyms appear in the English subtitled episode of any of the series, or an acronym is displayed in every other minute in a standard 40-minute-length English subtitled episode. We have selected the most frequent ones to analyze how three MT services (*Google Translate, DeepL* and *Yandex*) can handle their translation from English into Hungarian, comparing the results with the solutions provided by human translators. The three MT providers were selected based on their popularity (*Google Translate*), most recent positive reviews (*DeepL*) and non-English background (*Yandex*). Although there are many justified questions regarding the background of human translators (profile, native speakers or not, level of experience, number of translators involved in subtitling a season), unfortunately the answers are hardly available, and might range from enthusiastic amateurs (especially when no subtitler name is provided) to experienced, certified professionals working for *Netflix*, typically listing their names in the last subtitle line. As such, only renditions coming from identifiable native speakers of Hungarian will be considered, directing our attention to two *Netflix* series: *House of Cards* and *Designated Survivor*.

Preliminary findings show that acronyms can be handled by MT services with limited results; target language equivalents are offered for some of them (probably supported by extensive media use as well, such as *UN, DOD, LA* or *NYPD*), yet widely circulated (American) English acronyms are often preserved in the target alternative, constituting a real opportunity for human post-editing and proofreading (e.g., *EMP, ETA, FEMA* or *POTUS*). However, lack of consistency of human translators concerning the repetitive terms characteristic to TV series is also observed.

References

Barnett, A., & Doubleday, Z. (2021). Demonstrating the ascendancy of COVID-19 research using acronyms. *Scientometrics, 126*, 6127–6130.

Begg, C. B. (2017). Zero tolerance for acronyms. *Clinical Trials (London, England), 14*(6), 561–562.

Cannon, G. (1989). Abbreviations and Acronyms in English Word-Formation. *American Speech, 64*(2), 99–127.

Izura, C., & Playfoot, D. (2012). A normative study of acronyms and acronym naming. *Behavior Research Methods, 44*(3), 862–889.

Jamieson, E. (1968). Acronymania-A Modern Contagion. *New Scientist, 37*(29 February), 373–374.

Taghva, K., & Gilbreth, J. (1999). Recognizing acronyms and their definitions. *International Journal on Document Analysis and Recognition, 1999*(1), 191–198.

Yeates, S. A. (1999). Automatic Extraction of Acronyms from text. In D. Bainbridge & S. A. Yeates (Eds.), *The Third New Zealand Computer Science Research Students' Conference* (pp. 117–124). University of Waikato, Dept. of Computer Science.

Zahariev, M. (2004). A Linguistic Approach to Extracting Acronym Expansions from Text. *Knowledge and Information Systems, 2004*(6), 366–373.

••••

Tiffany Jandrain
Université de Mons-Hainaut, Belgium
tiffany.jandrain@umons.ac.be

Spoken versus Written: A Register Variation Analysis on the Use of the Passive Voice in Student Sight and Written Translations

This paper focuses on a register variation analysis that was carried out on the use of verbs in the passive voice in sight translations (i.e. the oral translations of a written text) and written translations made by Master students translating from English into French. In fact, studies have shown that students may find the transposition of register (i.e. "a language variety associated with both a particular situation of use and with pervasive linguistic features that serve important functions within that situation of use" (Biber & Conrad, 2009, p. 31)) from one language to another difficult because of the differences between English and French registers. For instance, Fawcett (1997), Gile (2005) and Vandaele (2015) have observed that even if students may often feel that language needs to be adapted according to register variables (the target audience, the text function, etc.), many fail to do this successfully in written translation.

This analysis was conducted on the passive voice of verbs, which "rearranges" the communication (Riegel et al., 2018, p. 666): the patient (affected by the action) is syntactically realised as the subject and becomes the "focused topic" of the sentence while the agent (instigating the action), if not completely removed, is "downgraded" (Van Hell et al., 2005, p. 244). In other words, the passive voice influences the organisation of the message and may be considered as a genre- or register-specific feature (Jisa et al., 2002; Van Hell et al., 2005). From this point of view, the analysis of the passive voice appears relevant to a register variation analysis.

Another factor that was taken into account in this analysis was the modes of speech, i.e. written and sight translations. In fact, since a register is determined by the mode in which a message is produced, this analysis aims at observing whether students used the passive voice in written and sight translations differently. It is important to emphasise that many consider that

sight translation "must 'sound' like the result of an oral communication and not like a written text" (Weber, 1990, p. 4).

1. The two research questions of this paper are the following:
2. To which extent does register have an influence on the use of passive voice in the students' translations?

Is the use of the passive voice different in written translations and in sight translations?

In order to answer those questions, two registers (a press article from *The Economist* (R1) and a popular science article from *New Scientist* (R2)) that are taught to Master students at University of Mons, Belgium, are compared, since a register analysis implies the analysis of at least two registers (Neumann, 2021). Four "multiple translation corpora" (i.e. corpora containing several translations of the same source texts (Granger & Lefer, 2020, p. 1186)) were thus compiled:

- a corpus of 14 written translations of R1
- a corpus of 17 sight translations of R1
- a corpus of 14 written translations of R2
- a corpus of 17 sight translations of R2.

The verbs used in the passive voice (VPV) in the translations were extracted by using the Corpus Query Language of the *Sketch Engine* concordancer (Kilgarriff et al., 2014). They were then analysed in the *R* software (R Core Team, 2020) (in the *RStudio* environment (Allaire, 2020)).

Results from the analysis of **written** translations in R1 and R2 show that there is a significant difference in active and passive voices according to registers (χ^2 (1) = 56.282, $p < 0.005$). Respectively, 51.1% and 100% of VPV in R1 and R2 are added by the students in their translations.

Results from the analysis of **sight** translations in R1 and R2 also show that there is a significant difference in active and passive voices according to registers (χ^2 (1) = 82.445, $p < 0.005$). Respectively, 52.2% and 100% of VPV in R1 and R2 are added by the students in their translations as well.

Results from the comparison between **written and sight** translations show that there is no significant difference in active and passive voices according to the mode of translation in both registers (χ^2 (1) = 0.43614, $p = 0.509$ in R1; χ^2 (1) = 0.00011411, $p = 0.9915$ in R2). In other words, the students' translations do not seem to be influenced by the mode of the message.

From a **qualitative** point of view, results show that 92.75% of VPV used in the sight translations are the same VPV used in the written translations in

R1, while none of the VPV used in the sight translations corresponds to the VPV used in the written translations in R2. Therefore, in our corpora, students seem to use the VPV in a more homogeneous way in their translations of the press article than in their translations of the popular science article.

References

Allaire, J. J. (2020). *RStudio* (1.3.1056) [Computer software]. RStudio Inc. https://rstudio.com/

Biber, D., & Conrad, S. (2009). *Register, genre, and style.* Cambridge: Cambridge University Press.

Fawcett, P. (1997). *Translation and language.* Manchester: St. Jerome Publishing.

Gile, D. (2005). *La traduction: La comprendre, l'apprendre.* Paris: Presses Universitaires de France.

Granger, S., & Lefer, M.-A. (2020). The Multilingual Student Translation corpus: A resource for translation teaching and research. *Language Resources and Evaluation, 54* (4), 1183–1199.

Jisa, H., Reilly, J. S., Verhoeven, L., Baruch, E., & Rosado, E. (2002). Passive voice constructions in written texts: A cross-linguistic developmental study. In R. A. Berman & L. Verhoeven (Eds.), *Written Language and Literacy* (Vol. 5). Amsterdam/Philadelphia: John Benjamins Publishing Company, 163-182.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1* (1), 7–36.

Neumann, S. (2021). Register and translation. In K. Mira, J. Munday, W. Zhenhua, & W. Pin (Eds.), *Systemic Functional Linguistics and Translation Studies.* London: Bloomsbury Academic, 65-82.

R Core Team. (2020). *R: A language and environment for statistical computing* (4.0.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Riegel, M., Pellat, J.-C., & Rioul, R. (2018). *Grammaire méthodique du français* (7th ed.). Paris: Presses Universitaires de Paris.

Van Hell, J. G., Verhoeven, L., Tak, M., & Van Oosterhout, M. (2005). To take a stance: A developmental study of the use of pronouns and passives in spoken and written narrative and expository texts in Dutch. *Journal of Pragmatics, 37* (2), 239–273.

Vandaele, S. (2015). La recherche traductologique dans les domaines de spécialité: Un nouveau tournant. *Meta, 60* (2), 209–235.

Weber, W. K. (1990). The Importance of Sight Translation in an Interpreter Training Program. In D. Bowen & M. Bowen (Eds.), *Interpreting. Yesterday, Today and Tomorrow*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 44-52.

••••

Marta Kajzer-Wietrzny
Adam Mickiewicz University, Poland
kajzer@amu.edu.pl

Dariusz Jakubowski
University of Silesia, Poland
jkbwsk@gmail.com

 Agnieszka Chmiel
Adam Mickiewicz University, Poland
achmiel@amu.edu.pl

Przemysław Janikowski
University of Silesia, Poland
przemyslaw.janikowski@us.edu.pl

Danijel Koržinek
Polish-Japanese Institute of Information Technology
danijel@pjwstk.edu.pl

Collocational patterns in interpretations into native and non-native language

Background

Aston (2018) argued that "the language of fluent interpreters relies heavily on recurrent formulaic phraseologies", and that their use may play a role in reducing the cognitive effort of simultaneous interpreters. Plevoets and Defrancq (2018) observed that "interpreters seem to have less processing difficulty not only with formulaic source texts, but also when their own production contains more formulaic sequences"

Looking at the frequency of recurrent bigrams (word combinations consisting of two elements) Kajzer-Wietrzny and Grabowski (2021) observed that interpreters use most frequent bigram types significantly more than original speakers, which points to the use of more formulaic language. However, examining the strength of collocations in intermodal analyses of mediated discourse, Ferraresi (2019) found that interpretations "display significantly lower collocational strength values" than speeches originally delivered in English, which suggests that interpretations are less formulaic.

Expanding Ferraresi's (2019) perspective, we investigate collocations in interpretations focusing, in particular, on comparing those performed into native and non-native tongue.

Despite the growing number of corpus studies on interpreting, little attention has been paid to the changes of the patterns of language use in interpreted discourse depending on directionality. Regardless of contrasting evidence concerning the preferred direction (L1-L2 or L2-L1) of interpreting, a number of studies outside of corpus interpreting studies highlights the complexity of the problem (Bartłomiejczyk, 2006, Janyan et al., 2009, García, 2015, Kim et al., 2018, Pokorn et al., 2020).

In our preliminary investigations we follow strictly the methodology proposed by Ferraresi (2019). Analysing collocations understood as sequences of words that occur together more often than predicted by chance (Jones and Sinclair 1974), we first calculate MI and T-score, i.e. association measures of all the collocations found in our corpus based on a large reference corpus of English (UkWaC; Baroni et al. 2009) and produce average scores for each text.

We focus on the use of frequent collocations of frequent words (higher T-score) and rarer collocations of greater collocational strength (higher MI-score). We also compare how many collocations used by native speakers and interpreters in our dataset do not occur in reference corpus. In the full paper we further intend to evaluate the impact of nativeness of the interpreter on the character of collocations in the target text with the use of regression modeling (with random effects for interpreter). Additionally, we plan to augment the analysis by including such extralinguistic aspects of interpreting and spoken communication as pauses or disfluencies.

Dataset

We used naturalistic data from the Polish Interpreting Corpus (Authors submitted). The PINC corpus comprises interpretations delivered at the European Parliament comprising around 200,000 tokens. As regards interpretations into English, PINC comprises interpretations delivered by Polish interpreters at the EP, i.e. interpretations into the active foreign language (B) covering about 40,000 tokens. PINC has recently been also enriched with a subcorpus (approximately 10,000 tokens) comprising interpretations into English as a mother tongue (A).

First results

Preliminary results show that native speakers (EN-Org), interpreters into foreign (EN-SI-B) and into native language (EN-SI-A) do not differ too much in their use of collocations.

Average MI-score per collocation in text is very similar across all varieties. Whether the speech was delivered impromptu or read out seems not to affect the average MI-score per collocation in texts produced by native speakers and interpreters working into native tongue. Interpreters working into foreign language seem to use rarer collocations comprising lower frequency words when interpreting texts originally read out.
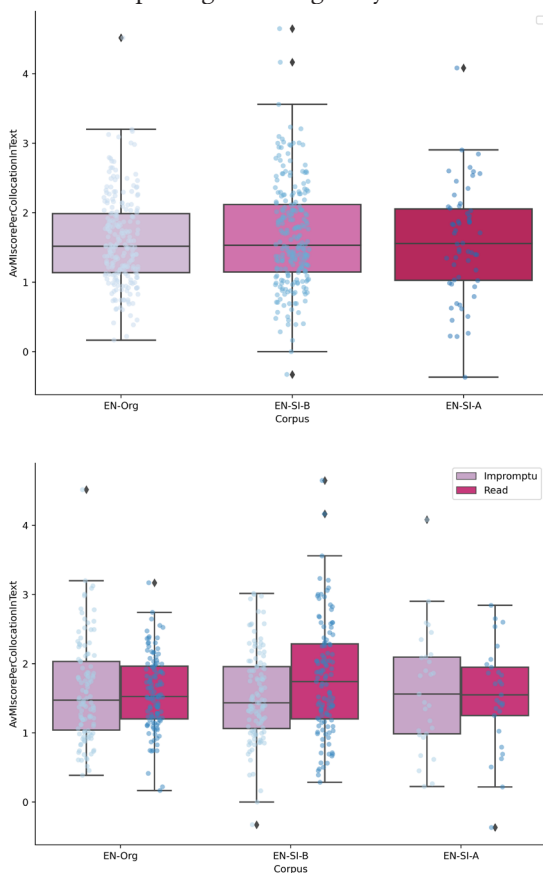
Texts produced by interpreters working into a foreign language display slightly higher average T-score per collocation, which means that they reach slightly more for frequent collocations composed of frequent words.

In the case of both interpreted varieties, average T-score per collocation in text increases when the original speech was read out, so with read out speeches interpreters tend to reach for more frequent collocations composed of more frequent words.
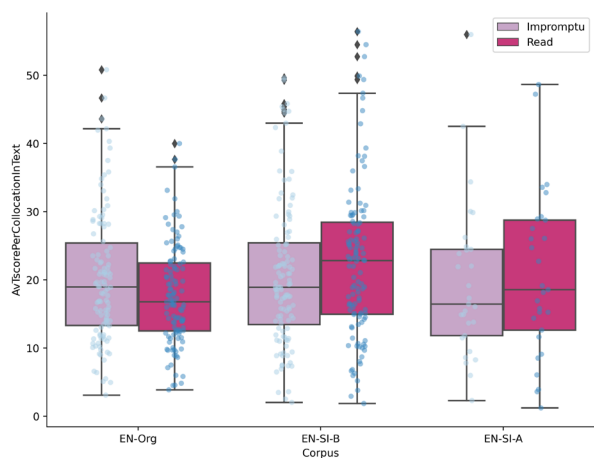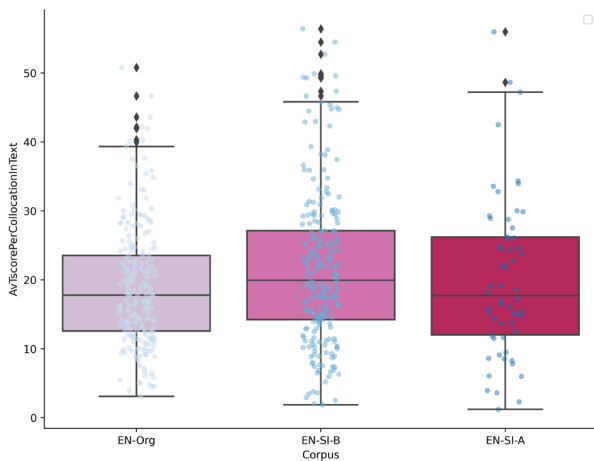
The number of collocations per text (normalized per 100 words) that do not occur in reference corpus is slightly lower in texts produced by interpreters working into a foreign language. The number of such collocations consistently increases when the speech was read out.
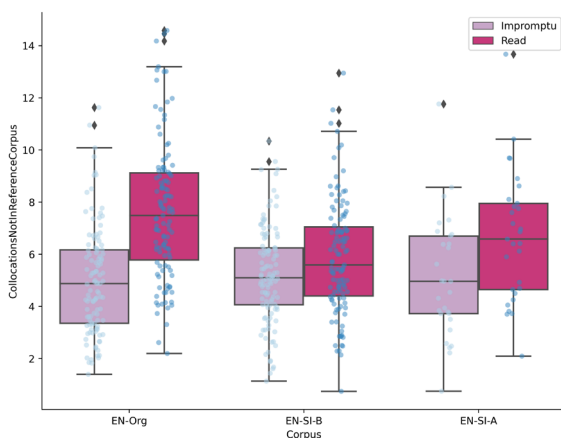
Preliminary conclusions

Comparing the output of interpreters working into their native and foreign language might shed light on how the additional constraint of producing a text in a foreign language affects the use of collocations. Assuming that formulae reduce the interpreters' effort (Aston 2018), it would be logical to expect that under "double constraint" of language mediation and bilingual production, interpreters' reliance on formulaic language would be stronger. The preliminary results show, however, that the differences between professional EU native and non-native interpreters might not be so striking and may be limited to the way they handle read out speeches.

References

Aston, G. (2018). "Acquiring the language of interpreters: A Corpus-based Approach". In: Russo, M, Bendazzoli, C. & Defrancq, B. (eds.), Making Way in Corpus-based Interpreting Studies. Singapore: Springer, 83–96.

Baroni, M, Bernardini, S., Ferraresi, A, Zanchetta, E. , "The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-crawled Corpora", Language Resources and Evaluation

Bartłomiejczyk, M. (2006). Strategies of simultaneous interpreting and directionality. *Interpreting* 8, 149–174. doi: 10.1075/intp.8.2.03bar

Ferraresi, A. (2019). Collocations in Contact: Exploring Constrained Varieties of English through Corpora. 2019. 203-221. 10.7370/93190.

Janyan, A., Popivanov, I., and Andonova, E. (2009). Concreteness effect and word cognate status: ERPs in single word translation. In: K. Alter, M. Horne, M. Lindgren, M. Roll, and J. von Koss Torkildsen (Eds.), *Brain Talk: Discourse With and in the Brain,* 21–30. Lund: Lunds Universitet.

Jones, S. R. H. & Sinclair, J. (1974). "English lexical collocations - A study in computational linguistics".

Kim, J. H., Kim, J. A., Lee, J. M., and Yang, J. H. (2018). Asymmetric cross-language activation of translations in Korean-English bilinguals. *J. Cogn. Sci.* 19, 1–30.

Plevoets, K, and Defrancq, B. "The Cognitive Load of Interpreters in the European Parliament. A Corpus-Based Study of Predictors for the Disfluency Uh(m)." *Interpreting*, vol. 20, no. 1, John Benjamins Publishing Company, 2018, 1–28, doi:10.1075/intp.00001.ple.

Pokorn, N. K., Blake, J., Reindl, D., and Peterlin, A. P. (2020). The influence of directionality on the quality of translation output in educational settings. *Interpret. Transl. Train.* 14, 58–78. doi: 10.1080/1750399X.2019.1594563

••••

Alina Karakanta
Fondazione Bruno Kessler / University of Trento
akarakanta@fbk.eu

Post-editing in automatic subtitling: auto-spotting in the spotlight

Subtitling is a multimodal phenomenon, where audio, image and text intermingle. Contrary to text translation, subtitling has additional requirements related to time (synchronisation and reading speed) and form (length and segmentation). With the increasing technologisation of subtitling to respond to the rising demand for multilingual subtitles, subtitling tools are being tailored to post-editing (PE). These solutions include not only automatic transcription and translation, but also automatic prediction of timestamps (auto-spotting). However, previous research in PE for subtitling has mainly focused on assessing quality and productivity when post-editing machine translated subtitles for which the timestamps are already provided (Volk et al. 2011, Etchegoyhen et al. 2014, Matusov et al. 2019, Koponen et al. 2020). Therefore, the effect of auto-spotting on the process of post-editing automatic subtitles (PEAS) has remained unexplored. In this work, we explore auto-spotting through a product- and process-based analysis of PEAS. Our research questions are as follows: What is the relationship between translation quality, auto-spotting accuracy and the time spent editing a subtitle? What is the contribution of editing the auto-spotting on the temporal and technical effort in PEAS?

Data

Three professional subtitlers, two for English→Italian, one for English→German, with experience in subtitling and PE post-edited 9 TED talks from the MuST-Cinema corpus (Karakanta et al., 2020), amounting to 1 hour of video (545 sentences/~10,000 words). The PE task was performed in MateSub[1], a novel free subtitling tool for post-editing fully automatic (translated, spotted and segmented) subtitles. Original and post-edited subtitles and productivity data (PE speed) were collected in process logs implemented in MateSub. Keystrokes were logged with InputLog (Leijten and Van Waes, 2013). We collected 3605 subtitles in total.

---

1   https://matesub.com/

Methods

For the first research question, we explore relationships between auto-spotting accuracy, translation quality and time spent on each subtitle. Based the process logs, we compute correlations between the following metrics:

Auto-spotting accuracy (dSpot): Total difference between the original position of an automatic subtitle and the final position of the post-edited subtitle (in seconds), computed as dSpot = abs(start_time$_{original}$ − start_time$_{final}$) + abs(end_time$_{original}$ − end_time$_{final}$).

Edit distance (ED): Changes in the translation of a subtitle are computed in terms of mean edit distance (Levenshtein) between the original and the post-edited subtitle text.

Post-editing speed (PES): Speed in words per minute. Minutes count the total time a subtitler remains "active" on a subtitle regardless of whether they perform an operation (type, drag, play) or not.

Since dSpot and ED are product-based metrics, we further conduct a process-based analysis of keystrokes of auto-spotting/text edits. The subtitlers adjust the auto-spotting by clicking and dragging subtitle boxes in the timeline at the bottom of the screen (see Figure 1). To isolate the spotting edits from other operations, we select mouse clicks inside the timeline area, based on x,y coordinates. We consider as text edits all keyboard operations, and mouse clicks in the subtitle area. We then compare the **total number of interaction events** and **total event duration** between the *spotting* and *text* groups.
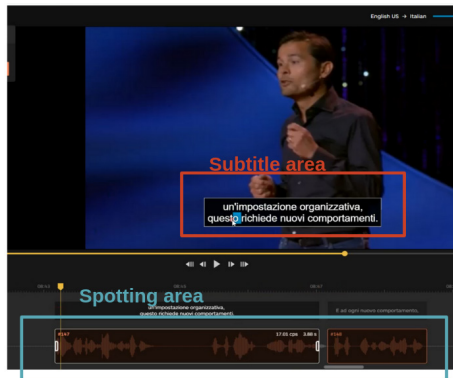
Figure 1. Subtitle-Spotting areas on the Matesub tool.

Results

Figure 2 shows a moderate negative correlation between PES-ED (-0.45) and a weak positive correlation between PES-dSpot (-0.26). The weak correlation suggests that a larger inaccuracy of auto-spotting does not necessarily imply larger temporal effort, while translation quality seems to be a more decisive factor. A stronger correlation was observed between ED-dSpot (0.64), confirming that changes in the translation will likely require adjustments of spotting, and vice versa.
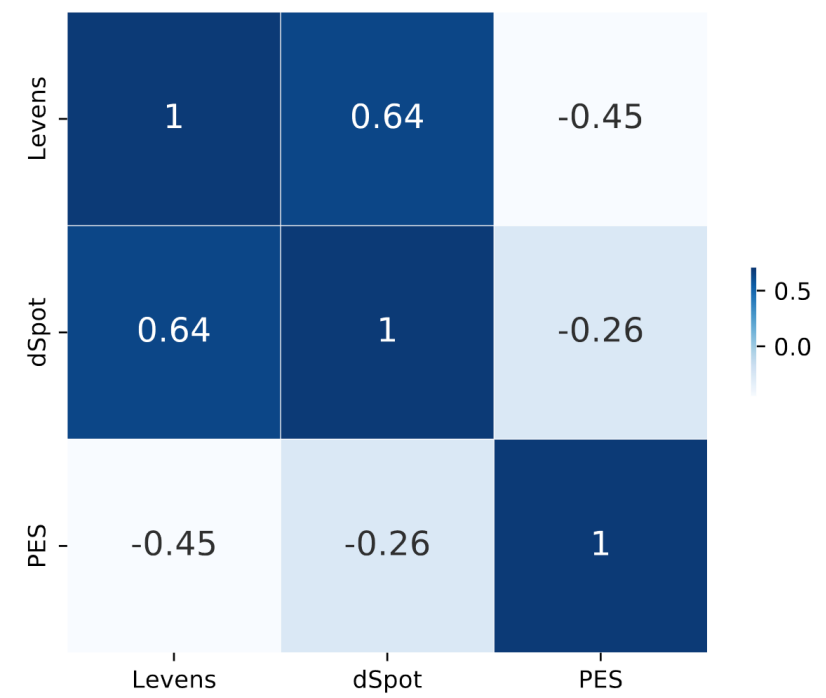
Figure 2. Pearson correlation between ED (Levens), dSpot and PES.

We then compare the technical and temporal effort of post-editing the text vs. the auto-spotting in terms of keystrokes (Table 1). The total number of events is much higher for text edits; however, spotting edits take 32-48% of the total PE duration. Despite the variability among subtitlers, spotting

edits take on average longer, since they require holding the mouse clicks to adjust the timestamps. Subtitlers perform several rounds of adjustments as they change the translation to satisfy subtitling constraints (e.g. reading speed). We also noticed that fine-grained adjustments can be more time-consuming than large adjustments. This is not captured by dSpot, as a product-based metric.

		Total events	Total event duration (mins)
It1	Spotting	10,459	134.77
	Text	44,029	199.08
It2	Spotting	8,516	116.93
	Text	54,329	121.26
De	Spotting	5,391	85.01
	Text	55,595	177.55

Table 1. Total interaction events and event duration for text and spotting edits.

Our preliminary analysis suggests that auto-spotting contributes to the technical and temporal effort in PEAS and is an important factor which should be considered both in evaluation of automatic subtitling and in subtitling process research. Based on this analysis, we will propose new automatic metrics for auto-spotting and the quality of automatic subtitling in general.

References

Thierry Etchegoyhen, Lindsay Bywood, Mark Fishel, Panayota Georgako-poulou, Jie Jiang, Gerard van Loenhout, Arantza del Pozo, Mirjam Se-pesy Maučec, Anja Turner, and Martin Volk. 2014. Machine Translati-on for Subtitling: A Large-Scale Evaluation. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*, pages 46–53.

Alina Karakanta, Matteo Negri, and Marco Turchi. 2020. MuST-Cinema: a Speech-to-Subtitles corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3727–3734, Marseille, France. European Language Resources Association.

Maarit Koponen, Umut Sulubacak, Kaisa Vitikainen, and Jörg Tiedemann. 2020b. MT for subtitling: User evaluation of post-editing productivity. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 115–124, Lisboa, Portugal. European Association for Machine Translation.

M. Leijten and L. Van Waes. 2013. Keystroke logging in writing research: Using inputlog to analyze writing processes. *Written Communication*, 30:358–392.

Evgeny Matusov, Patrick Wilken, and Yota Georgakopoulou. 2019. Customizing Neural Machine Translation for Subtitling. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 82–93, Florence, Italy. Association for Computational Linguistics.

Martin Volk, Rico Sennrich, Christian Hardmeier, and Frida Tidström. 2010. Machine Translation of TV Subtitles for Large Scale Production. In Ventsislav Zhechev, editor, *Proceedings of the Second Joint EM+/CNGL Workshop "Bringing MT to the User: Research on Integrating MT in the Translation Industry (JEC'10)*, pages 53–62, Denver, CO.

••••

Dorothy Kenny
Dublin City University, Ireland
dorothy.kenny@dcu.ie

Marion Winters
Heriot-Watt University
m.winters@hw.ac.uk

A text of one's own: post-editing and the affirmation of style

There is mounting interest in the use of machines to translate literary texts. Recent research has focused, for example, on the productivity and quality achievable by such means (e.g. Toral and Way 2018), on translators' perceptions of the process (Moorkens et al. 2018) and readers' perceptions of the product (Guerberof-Arenas and Toral 2020). This research, while valuable, is normally based on an experimental design that undermines ecological validity, and generally does not account for the way literary translators are likely to, or actually do, work with MT; nor can it be said to characterize published literary texts created on the basis of a machine translation. Authors (2020) take steps towards integrating the translator's voice in literary machine translation studies and enhancing the ecological validity of the experimental setting. The current study goes even further by investigating a full, published novel that was translated by machine and then post-edited by a literary translator. While the work was described by the translator as "painstaking re-translation" (Oeser 2020:22), the machine-translated and published post-edited versions of the novel nevertheless constitute a boon for the researchers, who can trace the exact edits made by the translator and analyze these edits from the point of view of their consistency with features of previous translations by the same translator. In short, they provide an opportunity to study whether the translator in question re-affirms his own style in post-editing the MT output, and thus whether he succeeds in making the text 'his own'.

The text at the heart of the research is Christopher Ishwerwood's *The World in the Evening*, its translation into German by DeepL (henceforth 'DeepL'; 104,969 tokens) and subsequent post-edited version by Oeser (henceforth 'OeserPE'; 103,974 tokens). The investigation of style is based in the first

instance on lexical techniques commonly used in stylometry in tasks such as authorship or translatorship attribution and in the comparison of translations (e.g. Rybicki 2012; Rybicki and Heydel 2013; Mastropierro 2018). The first, a positive keyword analysis, conducted using the keywords function in WordSmith 8.0, relies on the detection of unlemmatized types that occur with greater-than-expected frequency in the text under study (OeserPE) based on their frequency in the reference text (DeepL). Such words are considered keywords if they also pass all three tests of statistical significance used by WordSmith (Log Likelihood, Log Ratio and BIC). The relative frequency of positive keywords thus detected in OeserPE is subsequently compared with the same type's frequency in two larger corpora: a corpus of other translations by Oeser (just over 700,000 words; henceforth the 'Oeser Corpus') and a 3.5 million-word corpus of original German fiction, sourced through the *Institut für deutsche Sprache.* The analysis suggests strong 'marker words' (e.g. *weshalb* 'why', *stets* 'always', *all zu* 'too') that not only distinguish OeserPE from DeepL, but that prove to be indicative of Oeser's style over time, and to differentiate him from comparable writers in German. Evidence from interview data, suggests that such marker words are used consciously and are eminently explainable (Caballero et al. 2021) as a features of Oeser's style. Subsequent ongoing analysis using Burrows' Delta (Burrows 2002), which computes the mean of the absolute differences between the z-scores for each of the top 150 word forms in (1) the DeepL translation and the Oeser Corpus, and (2) OeserPE and the Oeser Corpus, also indicates that OeserPE "differs less" (Burrows 2002) from the Oeser Corpus than DeepL does, again indicating that as Oeser edits the DeepL output he, whether consciously or not, brings the text closer to his own style.

References

Burrows, J. (2002) 'Delta': A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3): 267–87.

Caballero, C., H. Calvo and I. Batyrshin, (2021) On Explainable Features for Translatorship Attribution: Unveiling the Translator's Style with Causality. *IEEE Access*, vol. 9, pp. 93195-93208, doi: 10.1109/ACCESS.2021.3093370.

Guerberof-Arenas, A. and A. Toral (2020) The Impact of Post-editing and Machine Translation on Creativity and Reading Experience. *Translation Spaces* 9(2): 255 – 282.

Kenny, D. and M. Winters (2020) Machine translation, ethics and the literary translator's voice. *Translation Spaces* 9(1):123-149.

Mastropierro, L. (2018) *Corpus Stylistics in Heart of Darkness and its Italian Translations.* London: Bloomsbury.

Moorkens, J., A. Toral, S. Castilho and A. Way (2018) Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces* 7(2): 240–262.

Oeser, H-C. (2020) Duel with DeepL. *Counterpoint* 4:19-23.

Rybicki, J. (2012) The Great Mystery of the (almost) Invisible Translator. Stylometry in Translation, in *Quantitative Methods in Corpus-Based Translation Studies,* Michael Oakes and Meng Ji (eds), Amsterdam/Philadelphia: Benjamins, 231-48.

Rybicki, J. and M. Heydel (2013) The stylistics and stylometry of collaborative translation: Woolf's Night and Day in Polish. *Literary and Linguistic Computing* 28(4): 708-717.

Toral, A, and A. Way (2018) What level of quality can Neural Machine Translation attain on literary text? in J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty (eds), in *Translation Quality Assessment: From Principles to Practice*, Heidelberg: Springer International Publishing, 263–287.

••••

Tímea Kovács
Károli Gáspár University of the Reformed Church in Hungary
kovacs.timea@kre.hu

A comparative analysis of neural machine- and human-translated and non-translated Hungarian and English legal texts

Owing to the exponential growth of neural machine translation, a paradigm shift has been witnessed regarding the role of translators and reviewers. As neural machine translation is increasingly more capable of modelling how natural languages work, the traditional tasks of translators are being gradually replaced by new challenges (Castilho et al., 2019). Consequently, more emphasis is placed on pre- and post-editing (revision) skills and competences (Pym, 2013; Robert et al., 2017), enabling the production of higher quality and near human-made translations. Therefore, the efficiency of pre- and post-editing largely depends on how aware translators are of the mechanisms and limitations of neural machine translation tools adopted in given language pairs (Lample et al., 2018).

There has been a myriad of scholars doing research on the peculiarities of English legal language use. Šarčevič (1997) and other scholars (Kjaer, 2007; Cao, 2012; Ződi, 2017) examined legal language use as to how legal texts can be classified as descriptive, perspective, or hybrid texts based on the theory of speech acts (Austin, 1962; Searle, 1979). Other scholars have focused on the grammatical and structural aspects of English legal language use (Pavlíčková, 2008; Bázlik et al., 2010). In addition to the pragmatic, grammatical, and structural features of legal language, legal translation has also received considerable attention. A significant development can also be observed in the field of presenting new research methods in legal translation and the possible applications of corpus linguistic tools (Biel & Engberg, 2013; Biel 2014, 2019; Khaydarova, 2019). There has been significant research into the use of Hungarian legal language (Dobos, 2014; Minya-Vinnai, 2018; Stíluskönyv, 2014; Tóth & Kurtán, 2017), as well. Nevertheless, there are relatively fewer academic publications on the translation of English-Hungarian and Hungarian-English legal texts (Balogh, 2020; Kovács, 2018, 2020).

In my presentation, I attempt to demonstrate through the qualitative (by means of Sketch Engine) and quantitative comparison of non-human (relying on neural machine translation applications, such as Microsoft Custom Translator and Google AutoML) translated legal texts to human-translated ones, the relevant challenges and dynamic contrasts arising in the process of translating. The presentation will focus on the use of use of adverbs, particularly *hereof* and *thereof*. The paper seeks to answer the question of which of the texts examined using corpus linguistic tools and complementary qualitative methods better corresponds to the language use trends observed in the source English text. Which target language text (human or machine translated) "gravitates" (Halverson 2003, 2010a, 2017) more towards the source language text. Based on the findings of research conducted previously by the author on the human and machine translations of the Hungarian Criminal Code and the Penal Code of California, it can be presumed that the frequency of the use of such adverbs in the human translated text follows more closely the patterns observed in the source language reference text. Nevertheless, the pattern of using such adverbs at higher, syntactic levels, in word combinations and sentences observed in the machine translated text converges more with the English reference texts (Kovács, 2021).

Through the qualitative and quantitative analysis of the Hungarian (source language) Fundamental Law of Hungary and its English (target language) machine and human translations as well as the (American) English (source language) text, particularly the U.S. Constitution, I aim to highlight the recurring patterns in the process of human and machine translation. I aim to examine the human- and machine-translated (Microsoft (Custom Translator) or Google (AutoML), texts of the Fundamental Law of Hungary and its human- and machine-translated texts. With a view to that, an ad-hoc corpus is built from the six texts and analysed through quantitative (corpus linguistic) and qualitative (manual) research methods.

Two non-translated texts, the English language U.S. Constitution and the Hungarian language Fundamental Law of Hungary will serve as monolingual source texts of human- and machine translation as well as source-language "reference texts". Their respective human- and machine-translated English and Hungarian texts will be analysed as parallel texts translated to a given target language.

I also aim to examine if the gravitational pull hypothesis described by Halverson (2003, 2010, 2017) in translated texts can be observed in my ad-

hoc corpora. The presentation is focused on the question whether human translated texts converge more towards the "presumed" characteristics of target language texts considered as reference texts, while the machine-translated text "gravitates" more towards the *actual* usage typical of target languages. The ultimate goal of my research is to examine what implications the growing prevalence of non-human (neural machine) translation has for language professionals.

References

Austin, J. L. (1962*). How to do things with words*. Clarendon Press: Oxford

B. Kovács, M. (1999). A funkcióigés szerkezetek a jogi szaknyelvben. *Magyar Nyelvőr* 123. évf., 388–94.

Balogh, D. (2020). *Műfajtudatosság a jogi szakfordításban és szakfordítóképzésben*. ELTE Doktori Iskola, unpublished dissertation
https://doi.org/10.15476/ELTE.2020.035

Bázlik, M. & Ambrus, P. & Beclawski, M. (2010). *The Grammatical Structure of Legal English*. Translegis.

Biel, Ł. (2014). The textual fit of translated EU law: a corpus-based study of deontic modality. *The Translator*, 20(3), 332–355.

Biel, L., & Engberg, J. (2013). Research models and methods in legal translation. *Linguistica Antverpiensia*, New Series–Themes in Translation Studies, (12).

Biel, Ł.; Engberg, J.; Ruano, R. M.; Sosoni, V. (eds.) (2019). *Research Methods in Legal Translation and Interpreting*. Routledge: London and New York.

Cao, D. (2012). *Legal translation. The encyclopedia of applied linguistics*.

Castilho, S., Gaspari, F., Moorkens, J. et al. (2019): Editors' foreword to the special issue on human factors in neural machine translation. *Machine Translation*. 33. 1–7. https://doi.org/10.1007/s10590-019-09231-y

[Last accessed: 15 January 2022].

Dobos, Cs. (2014). Lexikai átváltási műveletek a nyelven belüli fordítás- ban. *Alkalmazott Nyelvészeti Közlemények*, IX. évf., 2. szám, 63–80.

Halverson, S. (2003). The cognitive basis of translation universals, *Target,* 15(2), 197–241. https://doi.org/10.1075/target.15.2.02hal

Halverson, S. (2010). Cognitive Translation Studies and the merging of empirical paradigms: The case of 'literal translation', *Translation Spaces* 4(2), 310–340, https://doi.org/10.1075/ts.4.2.07hal

Halverson, S. (2017). Gravitational pull in translation. Testing a revised model. In De Sutter, G. & Lefer, M. A. & Delaere, I. (Eds.), *Empirical Translation Studies. New Methodological and Theoretical Traditions,* 9–47. De Gruyter Mouton. https://doi.org/10.1515/9783110459586

Khaydarova, U. (2019). Specific peculiarities of translation in legal documents. *Journal of Legal Studies and Research,* 5(5), 157–165.

Kjaer, A. L. (2007). Legal Translation in the European Union. A Research Field in Need of a New Approach. In Kredens, K., Goźdź-Roszkowski, S. (Eds.) *Language and the Law: International Outlooks,* 69–95. Frankfurt am Main: Peter Lang.

Kovács, T. (2018). Átváltási műveletek (megoldások) az angol-magyar jogi szaknyelvi fordításban. *Modern nyelvoktatás. Fordítástudomány különszám,* 25(2–3), 40–53.

Kovács, T. (2020). Interferenciajelenségek összehasonlító vizsgálata nem fordított magyar és angolról magyarra fordított jogi szaknyelvi szövegekben. In Szabóné Papp, J. (Ed.). *Alkalmazott Nyelvészeti Közlemények,* Interdiszciplináris tanulmányok, A Miskolci Egyetem Közleményei. Publications of the University of Miskolc. 13(2), 150–165. Miskolci Egyetemi Kiadó.

Kovács, T. (2021). Névmási szerkezetek kvantitatív és kvalitatív elemzése forrásnyelvi és fordított büntetőjogi szövegekben. *Modern Nyelvoktatás,* 2021(3–4), 58–72.

DOI: 10.51139/monye.2021.3-4.58.71

Lample, G., Ott, M., Conneau, A., Denoyer, L., Ranzato, M.A. (2018): *Phrase-based and Neural Unsupervised Machine Translation.* Cornell University, [online]. Available at: https://aclweb.org/anthology/D18-1549 [Last accessed 15 January 2022].

Minya, K., Vinnai, E. (2018). Hogyan írjunk érthetően? Kilendülés a jogi szaknyelv komfortzónájából. *Magyar jogi nyelv,* 1. szám, 13–18.

Pavlíčková, E. (2008). Ways of addressing in legal discourse. *Discourse and Interaction,* 1(1), 87–98. https://journals.muni.cz/discourse-and-interaction/article/view/6901/8267

Pym, A. (2013): Translation Skill-Sets in a Machine-Translation Age. *Meta: Translators' Journal,* 58(3), pp. 487–503.

Robert I.S., Terryn A.R., Ureel, J.J.J., and Remael A. (2017): Conceptualising translation revision competence: A pilot study on the 'tools and research' subcompetence. *The Journal of Specialised Translation,* 28, 293–316.

Šarčevič, S. (1997). *New Approach to Legal Translation*. Hague: Kluwer Law International.

TAUS. (2016). TAUS Post-Editing Guidelines, https://www.taus.net/think-tank/articles/postedit-articles/taus-post-editing-guidelines

Searle, J. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: University Press.

„*Stíluskönyv*": A joggyakorlat-elemző csoport összefoglaló véleménye a határozatszerkesztés tárgykörében 2014. évben végzett elemzés megállapításairól (2014). Kúria, 1–19. https://kuria-birosag.hu/sites/default/files/joggyak/osszefoglalo_1.pdf [Last accessed: 15 January 2022].

Tóth, J., Kurtán, Zs. (2017). A jogszabályok szövege – avagy a jogalkotók szövegelése. *Magyar Jogi Nyelv*, 1. szám, 34–36.

Ződi, Zs. (2017). Jogi szövegtípusok. *Magyar Jogi Nyelv*. (2), 20–30.

https://joginyelv.hu/jogi-szovegtipusok/

Sources:

Magyarország Alaptörvénye:

https://net.jogtar.hu/jogszabaly?docid=a1100425.atv.

[Last accessed: 15 January 2022].

The Fundamental Law of Hungary:

https://njt.hu/translated/doc/TheFundamentalLawofHungary_20191213_FIN.pdf.

[Last accessed: 15 January 2022].

U.S. Constitution:

archives.gov/founding-docs/constitution-transcript.

[Last accessed: 15 January 2022].

The human translation of the U.S. Constitution:

Bődy, P., Urbán, A. (eds.): *Szöveggyűjtemény az Amerikai Egyesült Államok történetéhez*, Budapest–Pécs, Dialóg Campus, 2001, 479–495.

••••

Raja Lahiani
UAE University
raja.lahiani@uaeu.ac.ae

Translate or Transliterate? When Metonymic Names are More than Proper Names

This study investigates problems surrounding translating and/or transliterating, and examines a case study and discusses how a poet/lover uses different appellations to purposefully address his beloved, which include the beloved's actual name and three different heteronyms that are examples of metonymy. As proper names transcend their identification functions in literature, they transmit knowledge, captivate the reader, provoke feelings, and also embody semantic, historical, social and geographical meanings. What makes a proper name problematic to Hermans is its "potential to acquire a semantic load which takes it beyond the 'singular' mode of signification of the proper name proper into the more 'general sphere of the common noun" (1988: 12). From here, Hermans links the translatability of proper names in "function of their 'semanticization'" and he includes with a "greater force" the proper names used in literary works. These works, to him, "show a greater concentration of 'motivated' or consciously 'loaded' names than non-literary texts" (1988: 13). Hermans classifies the proper names used in literature into "conventional" ("unmotivated") and "loaded," meaning "motivated" for they bear a semantic value and are thus "expressive" (1988: 13).

Hermans suggests four techniques to translate proper names: copying, transcribing, substituting and translating (1988: 13). Nyangeri and Wangari offer a relatively different framework that ascends from foreignizing strategies to domesticating ones: preservation, addition, omission, localization and creation (2019: 350). When it comes to Arabic on the one hand, and English and French on the other as the language combinations adopted in the experimental part of this study, copying would be out of context because of script differences. From Hermans' list, thus, the present work would adopt translation, substitution and transliteration. To these, two more techniques would be added based on Nyangeri and Wangari's model: interpolation (addition) and omission. A translator's choice of translation technique usually depends

on the communicative function(s) of the proper names and the translator's awareness of such functions.

This study considers the proper names used in a pre-Islamic ode by the poet Antara. The main theme of this poem is the poet/lover's desperate quest for his beloved's affection and his search for compensation in warfare and self-pride. Specific attention is given to verse lines 2, 5, 6, and 42. In these, Antara expresses his longing to his beloved by using different appellations, which reflects a relational complexity. The appellations traced in these verse lines are Abla, Umm al-Haytham, Ibnata Makhram, and Ibnata Mālik; the first is the beloved's conventional name while the rest are heteronyms. Contrary to the use of Abla, which is affectively unmarked and hence neutral, the use of heteronyms communicates affective markedness in the text. These heteronyms are instances of metonymy, a trope that occupies a considerable position in Arabic rhetoric. Metonymy "not only offers alternatives when naming but also opportunities for expressing nuance, giving emphasis and creating 'spin'" (Denroche, 2015: 5). Therefore, appellation variation and textual metonymy function in these verse lines not as elegant distinctions, but rather as functional stylistic choices for discourse purposes. The challenge of translating them rests in empowering the translation receiver to recognize the relation between the concepts in the same text unit and the function of their metonymic notions. While the original receiver easily understands that these names refer to the same woman and comprehends the appellation variations' appeals and functions, the translation receiver does not share the same range of expectations.

This study conducts a comparative critical assessment of a corpus consisting of 17 English and French translations that were produced between 1782 and 2020. Chronology and mode of discourse are taken into consideration. To highlight the improvement and influence, the earlier translations are mentioned first. Prose and verse translations are also evaluated separately with cross-references, wherever possible. In some cases, different modes of discourse may require distinct translation techniques. This study demonstrates how cultural and pragmatic losses are incurred within the process of correlating the verbal metonymic signs of the original culture to a different culture. The outcome is a misinterpretation of the source text's literariness and its pragmatic forces. As this study confirms, proper names are more than deictic symbols, and they bear functional communicative clues that determine specific translation techniques through which they can travel.

The translators faced the challenge of adopting a conscious decoding before rendering these names into the target language, where the problem is not simply a choice between translation, and transliteration. More fundamentally, they had to demonstrate enough awareness when deciding on the translation technique: interpolation, substitution, and omission being the main ones that they used. In the assessment process, the main weaknesses that were seen are implicit loss and modified loss. The comparative assessment reveals that most of the translators were not aware of the embedded cultural and idiosyncratic dimensions in the proper names in question and of the communicative clues that underlie their variations.

References

Denroche, C. (2015). *Metonymy and Language. A New Theory of Linguistic Processing.* New York and London: Routledge.

Hermans, T. (1988). On translating proper names, with reference to De Witte and Max Havelaar. In M. J. Wintle (ed.), *Modern Dutch Studies: Essays in Honour of Professor Peter King on the Occasion of his Retirement.* London: Bloomsbury, 11-24.

Nyangeri, N. and Wangari, R. (2019). Proper names in translation: Should they be translated or not? *Eastern African Literary and Cultural Studies*, 5 (3-4), 347-365.

••••

Ekaterina Lapshinova-Koltunski
Universität des Saarlandes, Germany
e.lapshinova@mx.uni-saarland.de

Heike Przybyl
Universität des Saarlandes, Germany
heike.przybyl@uni-saarland.de

Christina Pollkläsener
Universität des Saarlandes, Germany
chrpollk@gmail.com

Exploring Explicitation and Implicitation in Parallel Interpreting and Translation Corpora

The paper deals with the phenomena of implication and explicitation in translation and interpreting, one of the phenomena of translationese (see Gellerstam, 1986; Baker, 1993; Toury, 1995, amongst others), i.e. linguistic characteristics which distinguish translation from non-translation. Explicitation (Olohan and Baker, 2000; Blum-Kulka, 1986) is often related to the increased usage of discourse connectives. For instance, Gumul (2006, 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. Lapshinova-Koltunski et al. (2021) analysed discourse connectives with the help of neural semantic spaces to observe differences between translation and interpreting in terms of explicitation and implicitation effects – interpreting showed more implicitation than translation. The authors also stated the impact of relation type onto explicitation effect in translation (cognitively complex relations are not easily left out Blumenthal-Drame, 2021; Hoek et al., 2017) does not apply for interpreting.

Most corpus-based studies of explicitation and implicitation look into comparable data only, comparing distributions of discourse connectives in the subcorpora of sources, translations and comparable target originals. However, detection of explicitation and implicitation effects require analysis

of parallel data (see e.g. Marco, 2018; Becher, 2011). In this study, we look into parallel data to verify reported explicitation and implicitation trends in translation and interpreting. We analyse if discourse connectives in translation and interpreting have potential triggers in the sources, also paying attention to the degree of the explicitation signal. We inspect the the translational pairs of discourse connectives in the sources and in the targets to detect explicitation patterns, e.g. a weak signal connective in the source, e.g. *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 *but* in interpreting in example (1-c).

(1)
a. **Aber** *ich glaube, in einer Hinsicht gibt es Einigkeit...*(German source)
b. **However**, *I believe that in one respect there is consensus...*(English translation)
c. **but** *euh one thing we agree on...*(English 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 mode. We use the bidirectional English-German part of Europarl-UdS (Karakanta et al., 2018) and EPIC-UdS (Przybyl et al., 2021), which include officially published original speeches, as well as transcripts of the speeches delivered at the European Parliament aligned with their translations or interpretations, correspondingly. 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 Lapshinova-Koltunski et al. (2021) reported explicitation and implicitation effects observed in bilingual semantic spaces. Relying on connective lexicons and their occurrences in a reference corpus, 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. Then, we analyse the parallel to see (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/interpreter add items or change more general to more specific ones (explicitation). We 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. 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.

Frequency distributions of connectives (see visualisation for four connectives in Figure 1) show 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 intepreting, we observe more implicitation than equivalence as well as an unusual amount of explicitation for this connective (The differences are confirmed by a Pearson's Chi-squared test: *but, because* and *aber*: p-value < 2.2e-16, *weil*: p-value= 3.965e-08).

(a) but   (b) because   (c) aber   (d) weil
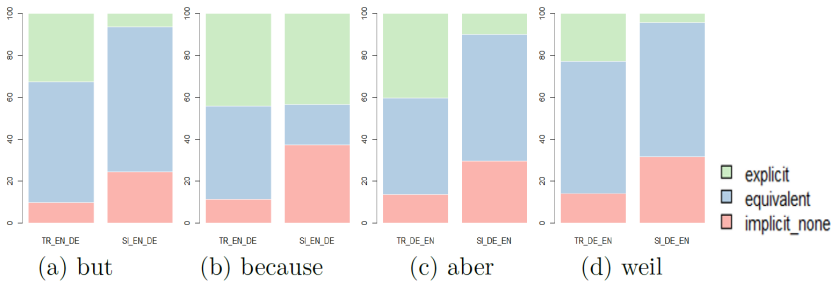
Figure 1: Distribution for translation/interpreting of *but*, *because*, *aber* and *weil*

In our presentation, we will show results on the distribution of connectives and their translations, as well as surprisal values indicating cognitive processing effort of translators/interpreters (source connectives) or readers/listeners (translation/interpreting)

References

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

Becher, V. (2011). Explicitation and implicitation in translation. A corpus-based study of English-German and German-English translations of business texts. PhD thesis, Staats-und Universit ätsbibliothek Hamburg Carl von Ossietzky.

Blum-Kulka, S. (1986). Shifts of Cohesion and Coherence in Translation. In House, J. and Blum-Kulka, S., editors, Interlingual and intercultural communication, pages 17–35. Gunter Narr, Tübingen.

Blumenthal-é, A. (2021). The online processing of causal and concessive relations: Comparing native speakers of english and german. Discourse Processes, 0(0):1–20.

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

Gumul, E. (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, 7:171–190.

Hoek, J., Zufferey, S., Evers-Vermeul, J., and Sanders, T. J. (2017). Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. Journal of Pragmatics, 121:113–131.

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

Karakanta, A., Vela, M., and Teich, E. (2018). Europarl-UdS: Preserving metadata from parliamentary debates. In Fišer, D., Eskevich, M., and de Jong, F., editors, Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA).

Lapshinova-Koltunski, E., Przybyl, H., and Bizzoni, Y. (2021). 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. ACL.

Marco, J. (2018). Connectives as indicators of explicitation in literary translationa study based on a comparable and parallel corpus. Target, 30:87–111.

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

Przybyl, H., Karakanta, A., Menzel, K., and Teich, E. (2022). Exploring linguistic variation in mediated discourse: translation vs. interpreting. In Kajzer-Wietrzny, M., Bernardini, S., Ferraresi, A., and Ivaska, I., editors, Empirical investigations into the forms of mediated discourse at the European Parliament, Translation and Multilingual Natural Language Processing. Language Science Press, Berlin.

Shlesinger, M. (1995). Shifts in cohesion in simultaneous interpreting. The Translator, 1:193–214.

Toury, G. (1995). Descriptive Translation Studies – and Beyond. John Benjamins, Amsterdam.

••••

Ekaterina Lapshinova-Koltunski
Saarland University
e.lapshinova@mx.uni-saarland.de

Maja Popović
ADAPT Centre
maja.popovic@adaptcentre.ie

Maarit Koponen
University of Eastern Finland
maarit.koponen@uef.fi

Differences in Translations of Amazon Reviews across Languages and Levels of Expertise

We analyse variation in human translations of Amazon reviews, as we find them interesting for a number of reasons: (a) they represent an informal register and would reveal variation in translator groups - Bizzoni and Lapshinova-Koltunski (2021) and Redelinghuys (2016) show that students are more repetitive in their linguistic choices than professionals; (b) they still remain challenging for machine translation - see an overview of studies on challenges posed by user-generated content (Al Sharou et al. 2021); (c) they reveal interesting cross-cultural differences. We rely on the existing works that report linguistic differences between professionals and students (Martínez and Teich, 2017; Redelinghuys, 2016; Bizzoni and Lapshinova-Koltunski, 2021; Kunilovskaya and Lapshinova-Koltunski, 2020; Lapshinova-Koltunski, 2020, De Sutter et al., 2017) and apply similar measures used by Popović (2020) for the analysis of transaltions by professionals, students and crowd workers.

Our parallel corpus (Lapshinova-Koltunski et al., 2022) contains English sources and their translations into Croatian, Russian and Finnish that were performed by both students and professionals, all native speakers of the corresponding languages. The number of translators varied from 4 to 20 per group, their experience - from 0 to 37 years. The sources include 196 Amazon product reviews (McAuley et al., 2015) on fourteen different topics with an equal number of positive and negative reviews. For each English texts, there are six translations: two per target language translated by professionals and students (see Table 1).

English source	Translations						total
	Croatian		Russian		Finnish		
	prof	stud	prof	stud	prof	stud	
15,236	14,003	13,940	14,233	14,247	11,709	12,213	95,581

Table 1: Corpus statistics.

We performed preliminary analysis of the data estimating overlaps and mismatches between two variants of translations for all the three languages under analysis. We calculated (1) word unigram matching, an F1-score, which indicates if different translators used the same words; and (2) edit distance (Levenshtein, 1966) indicating if different translators used different words in different order. We also use three additional metrics for mismatches which can be considered as a combination of the the two metrics above as presented by Popović and Ney (2011). They include (3a) word order mismatch – different translators used same words but in different positions, which indicates differences in the sentence structure; (3b) inflection mismatch – different translators used the same lemma but in different forms, which indicates morpho-syntactic differences, and (3c) lexical mismatch – different translators used different words (lemmas) and/or phrases, which indicates differences in lexical choice.

Overlap/distance between professional and student translations: hr=Croatian, ru=Russian and fi=Finnish.

measure	target language		
	hr	ru	fi
(1) word overlap↓	57.5	53.4	51.9
(2) edit distance↑	58.2	63.6	63.4
(3) mismatsches			
(a) word order↑	5.2	5.2	4.4
(b) inflection↑	7.0	6.2	10.3
(c) lexical↑	45.8	52.2	47.9

Table 2 shows that there is a variation in translations in our data. For instance, the word overlap (measured as F1 score) of 57.6 for Croatian means

that in each set of 100 words, 57.6 words are the same in professional and student translations, so the translators used the same words or forms of words. So, the higher the score, the more similar the translations are. In our data, Croatian student and professional translations seem to be most similar, and the Finnish translation most dissimilar. The edit distance of 58.2 (example of Croatian translations) between the same two texts should be interpreted as follows: in order to make the two translations identical, 58.2 of 100 words should be changed. The higher the score, the more dissimilar the student and professional translations are. In our data, the highest score is observed for Russian translations followed by the Finnish ones. Croatian translations seem to be more similar. The numbers for the three types of mismatches show that the greatest difference lie in lexical choices (45-55%), namely using different words. Different word forms are used in 6-9% cases, while same words in different order in 5-7.5% cases. In addition, we can see that variation between professional and student translations is dependent on the language pair. Finnish, as the morphologically richest language from the three, exhibits the highest difference in word form choice, and the lowest one in word order. As for lexical choice mismatch, it is highest for Russian. While further analysis is certainly required, one possible reason might be different transcriptions of named entities. We plan to annotate the data (lemmatise, part-of-speech tag and parse) to perform further analyses for a better understanding of these differences. We believe that the knowledge on variation in translation is valuable for studies on not only human but also machine translation, as it can provide a clearer understanding of paraphrases and translation quality. Overall, our preliminary analyses with shallow linguistic features show that there is variation in the translations produced by the different groups of translators, and these tendencies vary depending on the language pair involved.

References

Al Sharou, K., Li, Z., and Specia, L. (2021). Towards a better understanding of noise in natural language processing. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), pages 53–62.

Bizzoni, Y. and Lapshinova-Koltunski, E. (2021). Measuring translationese across levels of expertise: Are professionals more surprising than students? In Proceedings of the 23rd Nordic Conference on Computational Linguistics (No-

DaLiDa), pages 53–63, Reykjavik, Iceland (Online). Linköping University Electronic Press, Sweden.

De Sutter, G., Cappelle, B., De Clercq, O., Loock, R., and Plevoets, K. (2017). Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translations. Linguistica Antverpiensia, New Series–Themes in Translation Studies, 16:25–39.

Kunilovskaya, M. and Lapshinova-Koltunski, E. (2020). Lexicogrammatic translationese across two targets and competence levels. In Proceedings of LREC 2020, pages 4102–4112, Marseille, France.

Lapshinova-Koltunski, E. (2020). Exploring linguistic differences between novice and professional translators with text classification methods. In Vandevoorde, L., Daems, J., and Defrancq, B., editors, New Empirical Perspectives on Translation and Interpreting, Routledge Advances in Translation and Interpreting Studies, pages 215–238. Routledge.

Lapshinova-Koltunski, E., Popović, M., and Koponen, M. (2022). DiHuTra: a parallel corpus to analyse differences between human translations. In Proceedings of LREC 2022, Marseille, France.

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. Soviet Physics Doklady, 10:707–710.

Martínez, J. M. M. and Teich, E. (2017). Modeling routine in translation with entropy and surprisal: A comparison of learner and professional translations. In Cercel, L., Agnetta, M., and Lozano, M. T. A., editors, Kreativität und Hermeneutik in der Translation. Narr Francke Attempto Verlag.

McAuley, J., Targett, C., Shi, Q., and van den Hengel, A. (2015). Image-Based Recommendations on Styles and Substitutes. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2015), pages 43–52, Santiago, Chile.

Popović, M. (2020). On the differences between human translations. In Proceedings of the EAMT 2020, pages 365–374, Lisboa, Portugal.

Popović, M. and Ney, H. (2011). Towards automatic error analysis of machine translation output. Computational Linguistics, 37(4).

Redelinghuys, K. (2016). Levelling-out and register variation in the translations of experienced and inexperienced translators: a corpus-based study. Stellenbosch Papers in Linguistics, 45(0):189–220.

····

Marie-Aude Lefer
UCLouvain
marie-aude.lefer@uclouvain.be

Romane Bodart
UCLouvain
romane.bodart@uclouvain.be

Adam Obrusník
UCLouvain
adam.obrusnik@gmail.com

Justine Piette
UCLouvain
justine.piette@uclouvain.be

Post-editing quality assessment in translator education and beyond: Introducing the MTPEAS taxonomy

New professional profiles have emerged in recent years in the language industry (cf. Way 2020). One of them is that of post-editor (Ginovart Cid et al 2020). In academia, there have been numerous calls for the integration of machine translation post-editing (MTPE) specialized training into translation curricula (cf. Nitzke et al 2019). Concrete pedagogical proposals have been put forward in the last decade, mostly in the form of stand-alone modules dedicated to machine translation (MT) and MTPE (e.g. Guerberof & Moorkens 2019) or MTPE practical tasks (e.g. Pavlović & Antunović 2021). However, as rightly argued by Mellinger (2017), curriculum-wide initiatives are needed to fully support the acquisition of MTPE skills among translation students. A major hurdle for such a curriculum-wide implementation of MTPE across language-pair- and domain-specific practical courses is the assessment of students' productions. MTPE quality assessment (MTPEQA) in translator education is a complex process that involves three texts (the source text, the machine-translated text, and the post-edition) and, ultimately, boils down to the trainer's correction of students' edits (or lack thereof) of MT output. This is a time-consuming task, as trainers need to check that students have identified and appropriately corrected all MT errors, while

simultaneously checking that they have not introduced new errors while post-editing the MT. To date there are no widely-available, standardized MTPE annotation systems to be used in translator education. Incidentally, the same can be said about the language industry. While the issue of MTPE quality is frequently addressed by professional associations (see e.g. Gene & Guerrero 2022), MTPEQA is neglected by most stakeholders (cf. Doherty et al 2018), to the point that some language service providers explicitly acknowledge that it is not (yet) part of their standard workflows (this also holds for revision QA). To address this need in translator education (and beyond), we have devised a data-driven MTPE annotation system called MTPEAS (*Machine Translation Post-Editing Annotation System*). The MTPEAS taxonomy contains seven categories: Value-adding edits, Successful edits, Unnecessary edits, Incomplete edits, Error-introducing edits, Unsuccessful edits, and Missing edits. While Value-adding, Unnecessary and Error-introducing edits refer to changes made to error-free MT segments, the other categories in the taxonomy, namely Successful, Incomplete, Unsuccessful and Missing edits, are used to tag changes made to erroneous MT segments (or lack thereof, in the case of Missing edits). The seven MTPEAS categories are laid out in the form of a decision tree to guide MTPE trainers in their use of the system. MTPEAS is fully described in a manual (with definitions and authentic examples of all categories), which is available as an Open Educational Resource. With a view to offering high-quality, detailed feedback to students, MTPEAS is to be used in combination with the *Translation-oriented Annotation System* (TAS; Granger & Lefer 2021) developed within the *Multilingual Student Translation* project (Granger & Lefer 2020) to annotate errors in translations from scratch. The reason for relying on a system such as TAS, as opposed to translation quality assessment systems widely used in the translation industry, is that TAS has been primarily designed for pedagogical purposes, i.e. to support students in improving their translation and linguistic skills. The use of TAS makes it possible to identify the exact nature of the errors still present in the final MTPE products: Mechanics, Grammar and syntax, Lexis and terminology, Discourse and pragmatics, Register and style, Content transfer, Culture, and Brief. In the talk, we will introduce MTPEAS and present the results of an MTPEAS-based pilot study of 30 authentic 300-word English-to-French PE tasks performed by second-year master's students in translation. The PE tasks were carried out within the framework of two intensive courses in legal and financial translation. Stu-

dents who contributed data had all been previously trained in MT and MTPE through a dedicated course in the first year of their master's program. They were all familiar with the topics, terminology and text types at hand, having translated from scratch similar texts before completing the PE tasks. The MT contained approximately 50 errors, mostly related to Register and style, Lexis and terminology and Discourse and pragmatics. All student PE were annotated with MTPEAS and TAS by two raters working independently, who then solved all cases of disagreement (as to which segments to annotate and/or which MTPEAS and TAS categories to use). The final dataset contains 2000+ MTPEAS annotations, half of them complemented with TAS annotations. In our presentation we will provide first empirical insights into the quality of novice MTPE across the two specialized fields under scrutiny (legal and financial), laying particular emphasis on the issue of inter-rater agreement and how it makes it possible to assess the robustness of the MT-PEAS taxonomy we have developed.

References

Doherty, S., Moorkens, J., Gaspari, G. & Castilho, S. (2018). On education and training in translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari & S. Doherty (Eds.), *Translation Quality Assessment.* Cham: Springer, 95-106.

Gene, V. & Guerrero, L. (2022). *A Common Machine Translation Post-Editing Training Handbook for Academia, Clients, LSPs and Post-Editors.* GALA MTPE Training Special Interest Group.

Ginovart Cid, C., Colominas, C. & Oliver, A. (2020). Language industry views on the profile of the post-editor. *Translation Spaces*, 9 (2), 283-313.

Granger, S. & Lefer, M.-A. (2021). *Translation-oriented Annotation System manual (Version 2.0). CECL Papers 3.* Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain. Available online: https://uclouvain.be/en/research-institutes/ilc/cecl/cecl-papers.html

Granger, S. & Lefer, M.-A. (2020). The *Multilingual Student Translation corpus: A resource for translation teaching and research. Language Resources and Evaluation*, 54, 1183-1199.

Guerberof, A. & Moorkens, J. (2019). Machine translation and post-editing training as part of a master's programme. *The Journal of Specialised Translation*, 31, 217-238.

Mellinger, C.D. (2017). Translators and machine translation: knowledge and skills gaps in translator pedagogy. *The Interpreter and Translator Trainer*, 11 (4), 280-293.

Nitzke, J., Hansen-Schirra, S. & Canfora, C. (2019). Risk management and post-editing competence. *The Journal of Specialised Translation*, 31, 239-259.

Pavlović, N. & Antunović, G. (2021). Towards acquiring post-editing abilities through research-informed practical tasks. *Strani jezici: časopis za primijenjenu lingvistiku*, 50 (2), 185-205.

Way, C. (2020). Training and pedagogical implications. In E. Angelone, M. Ehrensberger-Dow & G. Massey (Eds.), *The Bloomsbury Companion to Language Industry Studies*. London: Bloomsbury Academic Press, 179-207.

••••

Simona Majhenič
University of Maribor, Slovenia
simona.majhenic@um.si

So, Discourse Markers – What do Interpreters do with Them?

One of the main goals of simultaneous interpretation (SI) is to facilitate communication between the source and target languages by conveying the speaker's message in unaltered form. As a survey among AIIC conference interpreters has shown, interpreters are very committed to this principle (Zwischenberger, 2013). One characteristic of SI is that it takes place under great time pressure. In turn, one would expect interpreters to focus more on the propositional and less on the non-propositional or metadiscoursive content such as discourse markers (DMs).

Nevertheless, although non-propositional, DMs carry meaning and modify the proposition (Fraser, 2009). In the example "Well now, you claim to be a good democrat," (Farage, 2012), the DMs "well now" contextualize the proposition. The problem is that due to the temporal constraints the interpreter has limited time to decipher the intended meaning of DMs (Defrancq et al., 2015), which are polysemous and multifunctional (Schiffrin, 2001) and require additional cognitive effort. As a result, DMs are occasionally disregarded or omitted, which in Farage's case means that the message will have a notably less negative connotation in the translation than in the original speech.

In light of the above, the following question arises: how does the ratio of DMs between the English source text (ST) and the Slovene target text (TT) compare? According to previous studies (Götz, 2017; 2020; Defrancq et al., 2015), the question cannot be adequately answered by a mere quantitative comparison of all identified DMs, as the TT can contain more DMs than the ST.

The present study, therefore, provides a systematic analysis of both omissions and additions of DMs in European Parliament speeches interpreted from English into Slovene and is the first to date to provide a general systematic overview of Slovenian DMs in SI. The speeches are part of the interpreting corpus TolAnSi, currently comprising manual transcripts equivalent to two hours of original and interpreted speeches. TolAnSi focuses on semi-

spontaneous and extemporaneous speeches by English native speakers. Since the setting is the European Parliament, all interpreters comply with accreditation standards, use the same equipment and are familiar with the speakers, allowing for the comparison of data in near laboratory conditions. The transcripts include disfluencies and truncated tokens. The multilingual corpus-based methodology builds on the classification of DMs according to Maschler (2009), who differentiates between textual (e.g., cause, consequence), interpersonal (e.g., agreement, maintaining contact), and cognitive DMs (e.g., realizing new information, or the need to rephrase). Maschler's (2009) three-pronged classification is especially suited for research into a mentally taxing process such as SI, as the domain of cognitive DMs encompasses DMs that occur during processes such as information processing.

DM types were annotated using Maschler's (2009) topology and the Praat annotation tool, and their corresponding translations were labelled as 'equivalent', 'non-equivalent', 'omitted', or 'added'. Next, the following hypotheses were tested: H1: Within the three domains, textual DMs were the least omitted group with respect to the ST; H2: Within the three domains, cognitive DMs were the most omitted group with respect to the ST; and H3: Within the three domains, cognitive DMs had the most additions with respect to the ST.

The results suggest that omissions and additions occur in all domains: however, omissions were the most frequent within the cognitive domain, where 58.49% of the identified DMs were omitted, and the least frequent within textual DMs, where 22.76% were omitted. Regarding the TT, 24.90% of all textual DMs were additions, while 78.74% of all cognitive DMs were added by the interpreter. Hence, all three hypotheses were confirmed. The results for H3 reflect the strain during SI due to the several efforts being juggled simultaneously by the interpreter (Gile, 2009), and provide grounds for further research into cognitive DMs as manifestations of potential translation difficulties the interpreter encounters. Moreover, the analysis of equivalent and non-equivalent translations of DMs in the corpus makes possible a quality assessment that focuses on the non-propositional or metadiscoursive items in SI, which, as the example above shows, can notably change the original message.

The findings indicate that DMs should not be systematically omitted, since doing so breaches one of the principles of SI, i.e., conveying the original message into the target language in integral form.

References

Defrancq, B., Plevoets, K., and Magnifico, C. (2015). Connective Items in Interpreting and Translation: Where Do They Come From? Springer International Publishing Switzerland. In J. Romero-Trillo (Eds.). *Yearbook of Corpus Linguistics and Pragmatics*, 3.

Farage, N. (2012). European Parliament Session 12. 9. 2012: https://www.europarl.europa.eu/plenary/en/vod.html?mode=unit&vodLanguage=EN&playerStartTime=20120912-12:32:27&playerEndTime=20120912-12:33:32#

Fraser, B. (2009). An account of discourse markers. *International Review of Pragmatics*, 1(2). 293–320.

Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Götz, A. (2017). Translating doubt: the case of the Hungarian discourse marker vajon. *Zooming Micro-Scale perspectives on cognition, translation and cross-cultural communication.* Peter Lang. 121–146.

Götz, A. (2020). Discourse markers and connectives in interpreted Hungarian discourse: A corpus-based investigation of discourse properties and their interdependence. *Speech Science* 1/2020. Hungarian Research Institute for Linguistics: Budapest. 259–284.

Maschler, Y. (2009). *Metalanguage in Interaction. Hebrew discourse markers.* John Benjamins Publishing Company. Amsterdam/Philadelphia.

Schiffrin, D. (2001). Discourse markers: language, meaning, and context. In D. Schiffrin, D. Tannen, and E. H. Hamilton (Eds.), *The handbook of discourse analysis.* Oxford: Blackwell, 54–75.

Zwischenberger, C. (2013). *Qualität und Rollenbilder beim simultanen Konferenzdolmetschen.* Bd. 1. Transkulturalität - Translation - Transfer. Berlin: Frank & Timme.

••••

Markéta Malá
Charles University, Prague
marketa.mala@pedf.cuni.cz

'Schodiště z mramoru' and 'the marble staircase': English divergent counterparts of Czech postmodifying prepositional phrases

The study examines the extent and types of correspondence between postmodifying prepositional phrases in Czech and their English translation counterparts. Dušková's (2015) research into syntactic constancy of clause elements has shown that not only is the syntactic function of a modifier preserved in about 90 per cent of noun phrases translated from Czech to English, but the form of the modification also remains largely constant. Given the fact that in English prepositional phrases are "by far the most common type of postmodification" (Biber et al. 1999: 606), the choice of a non-prepositional translation counterpart of a Czech prepositional postmodifier may appear quite surprising. Hasselgård's (2021) contrastive view of post-nominal prepositional phrases suggests that where divergent counterparts of prepositional postmodifiers do occur, they may reveal typological differences between the languages compared (cf. also Malá 2017). The present study sets out to explore the factors leading to the choice of such non-congruent counterparts.

The analysis relies on material drawn from the parallel corpus *InterCorp*. Four sub-corpora were used: two fiction and two non-fiction sub-corpora (Czech originals and their English translations, and Czech translations of English originals within each register). The Czech prepositions most frequently occurring in post-nominal position in the two registers were identified – *v*, *na*, *s*, *z*, *do*, *k* and *o*. Samples of the results of the query searching for the sequences 'noun + one of the prepositions' were checked manually to discard the chance sequences of a noun and adverbial. The English translation counterparts of the Czech postmodifying prepositional phrases were analysed, focussing on the non-prepositional correspondences. More specific queries were then formulated to further explore these types of counterparts, using the corpus as an "example bank" (Lüdeling & Kytö 2008: xiii). Apart from the typological differences between the predominantly synthetic

Czech and analytic English, the factors which may lead to the choice of a particular type of counterpart were also sought in the semantic relations between the head noun and the postmodifier, the syntactic structure of the noun phrase, the direction of translation, and register differences.

The English divergent counterparts of Czech postmodifying prepositional phrases were found to include premodification by nouns (ex. 1) or by *−ed* denominal adjectives (ex. 2); postmodification by a non-finite clause (exx 3, 4) or by a finite one (ex. 5), and other minor correspondences, such as premodifying adjectives, postmodifying adverbs or multi-word prepositions. Where the Czech head noun corresponds to an English verb form (often a gerund) the prepositional postmodification can be paralleled by an argument of the verb (ex. 6).

(1) *obchod s deskami* ('a shop with records') – *a record shop*
(2) *klobouk s pérem* ('a hat with a feather') – *a feathered hat*
(3) *sochy s lampama* ('statues with lamps') - *the statues bearing lamps*
(4) *prostředky k vybudování velké družiny* ('means for building a large retinue') – *the means to build a large retinue*
(5) *pokladnice na chodbě vedoucí k hlavnímu sálu* ('the treasury in the corridor leading to the main hall') - *the royal treasury, which is in the corridor leading to the main hall*
(6) *sňatek s lenochem* ('a marriage with a loafer') – *marrying a loafer*

As illustrated by exx 1-4 and 6, the divergent counterparts highlight the typological differences (and limitations) of the languages compared. In the most frequent type of divergent counterpart, premodification by a noun, the modifying function of the noun is indicated by its pre-head position in English, accentuating the grammatical function of the English word-order. In Czech, premodification would have to be expressed by an adjective, which may either be unavailable (ex. 1) or have a more general reference, cf. *zrcadlo v předsíni* ('the mirror in the front hall') vs. *předsíňové zrcadlo* ('a front-hall-ADJ mirror'). The English premodifying adjectives in *−ed* can be formed from noun phrases, corresponding to complex Czech postmodifiers, e.g. *tlustá dívka s kulatou tváří a copem* ('a fat girl with a round face and a pigtail') – *a fat, round-faced, pigtailed girl*. The postmodifying participles (e.g. *bearing*, *carrying*, *containing*, *wearing*) appear to serve as linking elements with quite general relational or posture meanings; the infinitives (particularly frequent

in non-fiction) signal the meaning of purpose or goal. The wider repertory of non-finite verb forms in English and a tendency to employ them to a much larger extent than in Czech can also be seen in the non-finite verbal counterparts of Czech (often de-verbal) nouns, with the prepositional postmodifier corresponding to an argument of the verb in English. Some counterparts (ex. 5), however, can hardly be accounted for by means of linguistic contrasts, and may rather be traced back to translation-inherent phenomena, such as a trend towards increased explicitness.

We hope to show that 'small words' (Hunston 2008), such as prepositions, can serve as a good starting point in contrastive research, drawing our attention to relatively 'large-scale' differences and similarities between the languages explored.

References

Biber, D. et al. (1999). *Longman Grammar of Spoken and Written English.* London: Longman.

Dušková, L. (2015). *From Syntax to Text: the Janus Face of Functional Sentence Perspective.* Praha: Karolinum.

Hasselgård, H. (2021). Lexicogrammar through colligation: Noun + Preposition in English and Norwegian. In A. Čermáková et al. (Eds.), *Crossing the Borders: Analysing Complex Contrastive Data. BeLLS* 11 (1), 145-161.

Hunston, S. (2008). Starting with the small words. Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics,* 13 (3), 271-295.

Lüdeling, A. & M. Kytö (Eds). (2008). *Corpus Linguistics: An International Handbook.* Volume 1. Berlin: Walter de Gryuter.

Malá, M. (2017). Non-prepositional English correspondences of Czech prepositional phrases: From function words to functional sentence perspective. In T. Egan & H. Dirdal (Eds.), *Cross-linguistic Correspondences. From Lexis to Genre.* Amsterdam: John Benjamins Publishing Company, 199-217.

Rosen, A., M. Vavřín & A.J. Zasina (2022). *InterCorp* – version 14 (January 17, 2022). Praha: Ústav Českého národního korpusu, FF UK. Available from WWW: https://kontext.korpus.cz/

••••

Fʀᴀɴçᴏɪs Mᴀɴɪᴇᴢ
CeRLA, Université Lyon 2
francois.maniez@univ-lyon2.fr

Mᴀʀíᴀ Bᴇʟéɴ Vɪʟʟᴀʀ Díᴀᴢ
CeRLA, Université Lyon 2
maria-belen.villar-diaz@univ-lyon2.fr

Sᴀɴᴅʀᴀ Gᴀʀʙᴀʀɪɴᴏ
CeRLA, Université Lyon 2
sandra.garbarino@univ-lyon2.fr

Use of English loanwords containing V-ING type forms in Spanish, French and Italian: a study based on the Subtitles 2018 and Leipzig corpora

Words starting with a verb root and ending with the *-ing* morpheme feature prominently among words borrowed from English in many Indo-European languages, and the rising popularity of the *-ing* morpheme has been attributed by some scholars (Picone 1996) to its nominalizing syntactic function. Such borrowings are frequently followed by the creation of equivalents coined by using native words in the receiving language, and occasionally by their inclusion in standard dictionaries (cf. the case of *brainstorming* and *remue-méninges* for French, Humbley 2008). Such neologisms occasionally present as hybrid borrowings (*surbooking* for *overbooking* in French) or pseudo-Anglicisms such as *mailing*.

Previous research (Maniez 2014) based on data from the multilingual Europarl corpus (Tiedemann 2009) showed that Italian borrowed more such forms than French (and French more than Spanish) in the European Parliament debates. With a view to ascertain whether comparable results would be obtained using data reflecting spoken language use, we queried the Subtitles 2018 corpus available on the Opus platform and studied the use of words beginning with a verb base and ending with the *-ing* morpheme (e.g. *shopping*) in Spanish, French and Italian.

In all three languages we found a strong tendency to borrow single-word terms but greater resistance to multi-word expressions, although some of them (*chewing gum*, *punching ball*, *sparring partner*) are occasionally used.

As was the case in the Europarl corpus, many loanwords concerned economy and finance (*rating, dumping, trading, holding*), as well as technology and communications (*mailing, roaming, streaming, zapping*), both other domains such as sports (*jogging*, *jumping*, *karting*, *kickboxing*, *rafting*) featured much more prominently in Subtitles 2018 than in Europarl.

All borrowings from English ending with the *–ing* morpheme were extracted from the Subtitles 2018 corpus in all three Romance languages and those which were used at least twenty times in one of the three languages (197 types) were analyzed quantitatively: Italian seemed to be the language with most such borrowings, while French showed relatively more resistance to *-ing* loanwords, and Spanish even more. Italian seemed to exhibit a lesser degree of resistance to *-ing* forms than Spanish and French for many such forms (*doping*, *overbooking*, *roaming*, *trading*), while some of them were actually used as translation equivalents for other English words (e.g. *mobbing* for *harassment*).

We then compared the figures we obtained for the most frequently used *-ing* forms in the Subtitles 2018 corpus with relative frequencies for those forms in a comparable corpus consisting of the large Web News corpora compiled by the University of Leipzig in the year 2020 (unlike the corpora available on the OPUS platform, the Leipzig corpora cannot be queried using wildcards in expressions such as < **ing*>, which meant that each of the selected verb forms was entered in separate queries in each of the three languages).

From a quantitative point of view, our results confirm the findings of previous research on the same topic based on the Europarl multilingual corpus. With only 87,5 *–ing* forms per million words (PMW), Spanish featured the lowest number of such borrowings, while French had over twice as many (211,4 PMW) and Italian almost four times as many (325,6 PMW). While Italian seems to have more *–ing* forms in most topical domains (sports, technology, business and finance), it also has very high numbers for many of the forms under study, with over 1,000 occurrences for ten of them (in descending frequency order: *streaming*, *screening*, *marketing*, *shopping*, *gaming*, *rating*, *feeling*, *pressing*, *meeting*, *sharing*).

References

ALVAR EZQUERRA M. (1995). *La formación de palabras en espagnol*, Madrid, Arco Libros.

BENARDI, R. L. L'italien des institutions publiques, une langue bien perméable aux anglicismes. *N 13–décembre 2014*, 16.

BISTARELLI, A. (2008). L'interferenza dell'inglese sull'italiano. *inTRAlinea, 10*, 1-11.

FURIASSI, C., PULCINI, V., et GONZÁLEZ, F. R. (ed.) (2012). *The anglicization of European lexis*. John Benjamins Publishing,.

GROSSMANN M., RAINER F. (dirs.) (*2004). La formazione delle parole in italiano*, Tübingen, Niemeyer.

HUMBLEY J. (2008). Emprunts, vrais et faux, dans le *Petit Robert 2007, In PRUVOST J. (dir.), Les journées des dictionnaires de Cergy : Dictionnaires et mots voyageurs. Les 40 ans du Petit Robert, de Paul Robert à Alain Rey*, 221-238, Herblay, Éditions des Silves.

MANIEZ, F. (2014). "Implantation of English terms including the -ING morpheme in French, Spanish and Italian: A corpus-based study of the debates of the European Parliament" in Pascaline DURY, José Carlos DE HOYOS, Julie MAKRI-MOREL, François MANIEZ, Vincent RENNER & María Belén VILLAR DÍAZ (dirs) *La néologie en langue de spécialité : détection, implantation et circulation des nouveaux termes*, pp. 189-201, Lyon, Travaux du CRTT.

PICONE M. D. (1996). *Anglicisms, Neologisms and Dynamic French*, Amsterdam, Benjamins.

RENNER, V., & FERNANDEZ-DOMINGUEZ, J. (2015). 6. False Anglicization in the Romance languages: A contrastive analysis of French, Spanish and Italian. In *Pseudo-English* (pp. 147-158). De Gruyter Mouton.

TIEDEMANN J. (2009). News from OPUS – A collection of multilingual parallel corpora with tools and interfaces, in NICOLOV N., ANGELOVA G., MITKOV R. (eds.), *Recent Advances in Natural Language Processing V: Selected papers from RANLP 2007*, 237-248, Amsterdam, Benjamins.

Internet sources

Leipzig Corpora Collection, https://corpora.uni-leipzig.de

OPUS, the open parallel corpus, http://opus.lingfil.uu.se/

••••

Josep Marco
Universitat Jaume I de Castellón, Spain
jmarco@uji.es

Ulrike Oster
Universitat Jaume I de Castellón, Spain
oster@uji.es

Passive construal as testing ground for the Gravitational Pull Hypothesis and Machine-translationese in the COVALT corpus

The Gravitational Pull Hypothesis (GPH) was put forward by Halverson (2003) as an attempt to provide a cognitive basis for alleged properties of translated text. On a more particular level, it also seeks to bring together two such properties that apparently contradict each other, namely over- and under-representation of target language (TL) typical features. It posits three potential causes of translational effects: patterns of salience or prototypicality (factor 1 – magnetism); conceptual structures/representation of the source language (SL) item (factor 2 – gravitational pull); and patterns of connectivity (factor 3 – connectivity).

Empirical research focusing on the GPH features studies by Halverson herself (e.g. 2017), Hareide (2017), Vandevoorde (2020) and De Sutter & Lefer (forthcoming). Within the framework of an ongoing research project, the COVALT group have been using several linguistic indicators to test out the GPH on a number of language combinations. One of these indicators is the category of constructions expressing passive construal. Three such constructions were identified as relatively salient in Spanish: the periphrastic passive, made up of the auxiliary verb *ser* ('to be') and the past participle of the main verb (e.g. *ser comprado* 'be bought'); the *se* construction with a non-specific agent, where the pronoun *se* is followed by a 3rd person singular or plural finite verb form (e.g. *se compra/compran* 'is/are bought'); and the 3rd person plural impersonal construction (e.g. *compran* 'they buy'). Oster and Tello's (2021) results lend support to the GPH, as it allowed to predict the differences between original and translated Spanish regarding these three passive constructions.

On the other hand, Literary Machine Translation (LMT) has recently emerged as a field of study in its own right. Literary translation has traditionally been seen as the ultimate frontier for MT because of the particular challenges posed by literary texts. But that is probably the reason why it has attracted the attention of MT scholars in the wake of the development of neural MT systems. Toral (2019) argued that, just as translated language differs from non-translated language, a parallel claim could be made for post-edited machine-translated language, which he calls *post-editese*. In fact, post-editese would be an exacerbated kind of translationese. Marco (2021) claimed that it is still possible to go one step further and regard *machine-translationese* as an exacerbated form of post-editese. This claim was borne out by his data.

In this paper we aim to test out Toral's MT exacerbation effect on the language combinations German-Spanish and English-Catalan. For each of these combinations, a three-component parallel corpus is available, made up of source texts (ST), human translations (HT) and machine translations (MT). The translation effects found by Oster and Tello (2021) have been used as the starting-point for the following predictions:

The *ser* + past participle construction will be over-represented in both EN-CA and DE-ES (though less so in the latter) due to gravitational pull.

The *se* construction will be over-represented in both EN-CA and DE-ES due to magnetism.

The 3rd person plural impersonal active will be under-represented in both EN-CA and DE-ES due to low connectivity between this TL construction and its SL triggers.

A preliminary analysis of a random sample of 250 query matches for each construction in TL non-translations, HTs and MTs (for each language combination) yields the results shown in Tables 1 and 2.

	CA	freq dist	HT	freq dist	MT	freq dist
ser + PP	586,99	12,60	793,04	38,47	948,88	34,27
es/se + 3rd p sing/pl	1694,58	36,39	984,40	47,75	1105,57	39,93
3rd p plural	2375,60	51,01	284,04	13,78	714,28	25,80
	4657,17		2061,49		2768,73	

Table 1. Preliminary results for the English-Catalan sub-corpus.

	ES	Rel f	HT	Rel f	MT	Rel f
ser + PP	536.7	20.4%	474.1	19.7%	732.4	26.1%
se + 3rd p sing/pl	997.7	37.9%	1,470.1	61.2%	1,758.0	62.6%
3rd p plural	1,096.2	41.7%	456.9	19.0%	315.7	11.3%
TOTAL	2,630.6		2,401.1		2,806.2	

Table 2. Preliminary results for the German-Spanish sub-corpus.

These results confirm our predictions for the three constructions in German-Spanish, but not in English-Catalan, where results for MTs stand somewhere in between non-translations and HTs and no exacerbation effect, therefore, is observed. Full analysis will include a larger sample of the three constructions in the six corpus components under scrutiny as well as an analysis of the four parallel sub-corpora (HTs and MTs in both language pairs) in order to identify the main ST triggers of the passive constructions and determine the degree of connectivity between ST triggers and TT solutions.

References

Halverson, Sandra (2003). The cognitive basis of translation universals. *Target*, 15 (2), 197-241.

Halverson, Sandra (2017). Developing a cognitive semantic model: magnetism, gravitational pull, and questions of data and method. In Gert de Sutter, Marie-Aude Lefer & Isabelle Delaere (Eds.), *Empirical Translation Studies. New Methods and Theoretical Traditions*. Berlin: Mouton de Gruyter, 9-45.

Hareide, Lidun (2017). Is there gravitational pull in translation? A corpus-based test of the *Gravitational Pull Hypothesis on the language pairs Norwegian-Spanish and English-Spanish. In Meng Ji, Michael Oakes, Li Defeng & Lidun Hareide (Eds.), Corpus Methodologies Explained. An Empirical Approach to Translation Studies*. London/New York: Routledge, 188-231.

Lefer, Marie-Aude & Gert De Sutter. Forthcoming. Using the Gravitational Pull Hypothesis to explain patterns in interpreting and translation: The

case of concatenated nouns in mediated European Parliament discourse. In Marta Kajzer-Wietrzny, Adriano Ferraresi, Ilmari Ivaska & Silvia Bernardini (Eds.), *Empirical investigations into the forms of mediated discourse at the European Parliament*. Berlin: Language Science Press.

Marco, Josep (2021). Machine-translationese in the English-Catalan sub-corpus of COVALT. Paper presented at the *PETRA-E Conference 2021: Literary Translation Studies Today & Tomorrow*, Trinity College Dublin, 4-5 November 2021.

Oster, Ulrike & Isabel Tello (2021). Torn between source language constructions and target language expectations. Translating passive construal. Paper presented at *UCCTS 2021 | Using Corpora in Contrastive and Translation Studies 6th edition)*, Bertinoro, 9-11 September 2021.

Toral, Antonio (2019). Post-editese: an Exacerbated Translationese. *arXiv preprint arXiv:1907.00900.*

Vandevoorde, Lore (2020). *Semantic differences in translation: Exploring the field of inchoativity*. Berlin: Language Science Press.

••••

JESSE MARION
Université de Mons-Hainaut, Belgium
jesse.marion@umons.ac.be

Translating Quantity Modification: A Case Study of English *all* and French *tout*

Although quantifiers themselves have been studied thoroughly (e.g. Barwise & Cooper 1981; Gärdenfors 1987; Langacker 1991, 2008, 2016, 2017; Doetjes 1997; Benninger 1999; Radden & Dirven 2007), the modification of quantifiers, i.e. quantity modification, has received scant attention, with the exception of Njende et al. (2015, 2017) and Davidse et al. (2018, fc.). Examples (1) and (2) below illustrate how two relative quantifiers, a subtype of quantifiers comparing a predicated mass P expressed by the quantifier to a reference mass R_T (Langacker 1991), undergo quantity modification in English and French respectively, i.e. how the quantity they denote is altered by a modifier highlighting the partial (as in (1) and (2)), full or non-coincidence between P and R_T.

1. *Almost* **all** of my friends have been fined either for having no seatbelt, or having no lights on their bikes. (YCCQA_uk)
2. Beaucoup font aussi la lecture de video, comme Vlc qui peut lire *presque* **tous** les formats video existant sans installation de codec. (YCCQA_fr)
   'Many read videos too, such as VLC, which can read *almost* **all** existing video formats without installing codecs.' (Translation mine)

This corpus-based synchronic study investigates quantity modification in translation. More precisely, it aims to inventory the choices made and translation strategies used by translators encountering modified quantifiers.

English *all* and French *tout* were selected for this study as they can be modified by a variety of adverbs and are relatively frequent in various types of speech. These two relative quantifiers can pair with proportional modifiers (Njende et al. 2017, building on Paradis 1997, 2000, 2001) to further detail the degree of coincidence between a predicated mass P and a reference mass R_T, and can undergo three types of modification: totality modification (3), which flags full coincidence of P and R_T (Njende et al. 2017); approximat-

ing modification (4), which highlights partial coincidence of P vis-à-vis R_T (Njende et al. 2017), and negational modification (5), which stresses the non-coincidence between P and R_T by means of a negator serving as a modifier of the quantifier (Marion 2021).

As the French negative construction *(ne)...pas* can take up different positions in the sentence, it will also prove interesting to study whether negational modification in English is kept in the French translations or whether it is replaced by, for example, sentential negation.

A semantic specification phenomenon (SSP, underlined in (5)) was often attested in the monolingual data when a negational modifier preceded a quantifier. SSPs consist in the addition of an element specifying the semantics of negational modification by means of setting up an explicit contrastive relationship (e.g. by adding *many* after *not all*, as in (5)) or by means of exemplifying (e.g. listing possibilities illustrating how P is different from R_T). This study will verify whether SSPs also occur in translation and, if so, whether they are equally frequent, and account for the way translators handle SSPs. The results obtained will allow to better understand the vague semantics of negational modifiers and will contribute to expanding the theory on quantity modification.

3. *Absolutely* **all** the expenses (…).[1]
4. (…) allergic to *pretty much* **all** animal hair.[1]
5. *Not* **all** people who don't vote for Obama are racist. Actually, many of them aren't.[1]

Data were extracted from the sixth version of the Europarl Direct parallel corpus (Cartoni & Meyer 2012), favoured over the seventh version as the former benefits from accurate language tagging. Europarl Direct consists of proceedings from the European parliament and totals approximately 50 million words per language collected from April 1996 to December 2010 (Koehn 2005, Cartoni & Meyer 2012). Provided that sufficient data was available, datasets of 300 instances per language pair were extracted using AntConc (Anthony 2010) as the source files were already aligned, thus making a parallel concordancer unnecessary. On the basis of the data selected, a parallel study contrasting the translations of English *all* into French and French *tout* into English was performed.

Preliminary results suggest translators either opt for quite literal, formally corresponding (Catford 1965) translations, as in (6), or resort to a more synthetic translation, as in (7) where the combination *almost all* is rendered as the compound noun *quasi-totalité* changing the quantity modification into a type of degree modification.

6. *Virtually* **all** our states (...). (Europarl6_uk)
    '*Quasiment* **tous** nos États (...).' (Europarl6_fr)
7. (...) for *almost* **all** its exports (...). (Europarl6_uk)
    '(...) pour la ***quasi-totalité*** de ses exportations (...).' (Europarl6_fr)

Not only will this study allow to explore quantity modification in translation and chart hitherto unexplored territory, but it will also shed light on the possible translations for both items under scrutiny and the main strategies translators resort to when confronted with the quantity modification of *all* and *tout*. A better understanding of how negational modification works and is translated into both languages under study is also targeted in this study.

References

Book

Catford, J. C. (1965). *A Linguistic Theory of Translation.* Oxford University Press.

Doetjes, J. S. (1997). *Quantifiers and selection: On the distribution of quantifying expressions in French, Dutch and English.* Holland Academic Graphics.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar* (Vol. 2, Descriptive Application). Stanford University Press.

Langacker, R. W. (2008). *Cognitive Grammar – A Basic Introduction.* Oxford University Press.

Munday, J. (2016). *Introducing Translation Studies* (4th ed.). Routledge.

Paradis, C. (1997). *Degree modifiers of adjectives in spoken British English.* Lund University Press.

Radden, G., & Dirven, R. (2007). *Cognitive English Grammar* (Vol. 2, Cognitive Linguistics in Practice). Benjamins.

Article in edited volume

Paradis, C. (2000). Reinforcing adjectives: A cognitive semantic perspective on grammaticalization. In R. Bermudez-Otero, D. Denison, R. Hogg & C.

B. McCully (Eds.), *Generative theory and corpus studies* (Topics in English linguistics *31*), 233- 258). Mouton de Gruyter.

Journal article

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. Linguistics and Philosophy. *Linguistics and Philosophy*, 4 (2), 159-219.

Gärdenfors, P. (Ed.) (1987). Generalized Quantifiers. *Linguistic and Logical Approaches*, 31. D. Reidel Publishing Company.

Langacker, R. W. (2016). Nominal grounding and English quantifiers. *Cognitive Linguistic Studies*, 3 (1), 1-31. https://doi.org/10.1075/cogls.3.1.01lan

Langacker, R. W. (2017). Grounding, semantic functions, and absolute quantifiers. *English Text Constructions*, 10 (2), 233-248. https://doi.org/10.1075/etc.10.2.03lan

Njende, N. M., Davidse, K., & Ghesquière, L. (2017). Precious few and practically all: A cognitive grammar approach to the modification of quantifiers. *Leuven Working Papers in Linguistics*, 34, 96-121.

Paradis, C. (2001). Adjectives and boundedness. *Cognitive Linguistics*, 12 (1), 47-65.

Presentation

Davidse, K., Ghesquière, L., & Njende, N. M. (2018). *Pursuing the analogy between quality and quantity modification* [PowerPoint presentation]. BAAHE conference, Mons, Belgium.

Marion, J. (2021). *Almost* all *and presque* tout*: A corpus-based study of quantity modification with English* all *and French* tout [Paper presentation]. Using Corpora in Contrastive and Translation Studies (6th edition), Bertinoro, Italy.

Njende, N. M., Ghesquière, L., & Davidse, K. (2015). *Re-assessing the absolute quantification restriction in existential constructions* [Paper presentation]. 48th Annual Meeting of the Societas Linguistica Europaea (SLE 48), Leiden, the Netherlands.

Corpus & software

Anthony, L. (2010). *AntConc* (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. www.antlab.sci.waseda.ac.jp.

Cartoni, B., & Meyer, T. (2012). Extracting Directional and Comparable Corpora from a Multilingual Corpus for Translation Studies. *Proceedings 8th International Conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey. Available online at http://www.lrec-conf.org/proceedings/lrec2012/pdf/188_Paper.pdf.

De Smet, H. (2009). *Yahoo-based Contrastive Corpus of Questions and Answers (YCCQA).* varieng.helsinki.fi/CoRD/corpora/YCCQA.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *MT Summit*, *5*, 79-86. https://aclanthology.org/2005.mtsummit-papers.11.pdf

Doctoral dissertation

Benninger, C. (1999). *De la quantité aux substantifs quantificateurs* [Doctoral dissertation]. Metz : University of Metz. Available online at www.gallica.bnf.fr/ark:/12148/bpt6k33703112

••••

Kayo Matsushita
Rikkyo University
kayo.matsushita@rikkyo.ac.jp

Masaru Yamada
Rikkyo University
masaru.yamada@rikkyo.ac.jp

Towards the establishment of a quality assessment framework for interpreting performance

In recent years, the assessment of translation quality, whether for the output of human translation or machine translation, has been carried out under a common quality evaluation framework such as MQM (www.qt21.eu). This is partly due to machine translation quality having surpassed the level of quality that can be measured by automatic evaluation such as BLEU. On the other hand, subjective evaluations by humans have not always been as reliable as possible (Freitag et al., 2021). This is why researchers are beginning to refer to a common translation quality assessment framework such as MQM.

So far, no similar movement has been observed in the field of interpreting. Even though the issue of quality has been a central topic in interpreting studies, there has not been concrete movement towards establishing a quality assessment framework that focuses specifically on the output of human interpretation rather than on interpreter competence. Several reasons have been identified (see Collados Aís & Garcia Becerra, 2015 for an overview), including the different perspectives of potential evaluators (e.g., interpreters, clients, end-users, speakers) and the ungeneralizable nature of interpreter performance which continuously responds to changing situational variables and unstable source utterances. All these factors make the quality assessment of interpreting performance much more complex and challenging compared to that of translation.

However, technological advancement driven by artificial intelligence (AI), already changing the game in the translation industry, is starting to impact the interpreting industry as well. Global companies such as Google, Baidu, and Microsoft, as well as researchers from around the world, are working on the development of machine interpretation (or speech-to-speech translation) technology, showing concrete signs of progress (e.g., Kitagawa, 2020). This

brings about an emerging need to compare and analyze human and machine performance, which calls for a quality framework that utilizes observable and measurable indicators.

In order to explore the possibility of establishing a standardized framework, this study assessed the quality of interpreting performances by ten professional interpreters in a pseudo-authentic setting. Using the cloud-based video conferencing system Zoom, the professionals interpreted a videorecorded press conference given by then-Australian Foreign Minister Julie Bishop in 2016. Remote Simultaneous Interpreting (RSI) has become the central means of providing interpreting services because of the COVID-19 pandemic (Matsushita, 2020). Not only did the use of RSI allow us to replicate the   working environment of interpreters today, but it also enabled us to minimize the difference between a live press conference and a pre-recorded one since they look identical when streamed through Zoom. The interpreters performed interpreting between English and Japanese, and their performances were recorded, transcribed, aligned, and annotated.

For the assessment method, this study selected key indicators of interpreting quality based on findings from prior research (e.g., Bühler, 1986; Moser-Mercer, 1996; Zwischenberger, 2010) and the metrics used in MQM for translation quality assessment. Since the purpose of this study is to find observable and measurable quality indicators unique to interpreting, we chose "fluency" and "synchronicity" as indicators of the quality of the interpreter's delivery. Fluency was calculated both by the volume of output and pause frequency and duration. A third indicator, "correct terminology," was added to evaluate content-related quality in a quantifiable manner. MQM's error category of "inconsistent with domain" was used to identify the number of cases where the interpreter failed to accurately render pre-selected terms commonly used in diplomacy and global politics.

We then assessed the quality of the interpreter performances by using the selected indicators and comparing the results with the actual onsite performance recorded and included in the preexisting corpus of interpreter-mediated press conferences (the JNPC Corpus) to look for any notable differences between online and onsite interpretations that could potentially affect quality. We then correlated the selected quality indicators and the results with quality assessments performed manually by human evaluators using a rubric based on Bühler's (1986) widely used criteria. Initial analysis revealed that the criteria selected for this study have the potential to assess the qual-

ity of interpreting performance to a certain extent. The results also showed a moderate correlation with existing subjective assessment methods currently used in the interpreting industry.

It is hoped that this study will lay the foundation for more discussions on interpreting performance quality and for various approaches towards establishing a quality assessment framework for interpreting practices to emerge

References

Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua,* 5(4), 231–235.

Collados Aís, A., & Garcia Becerra, O. (2015). Quality. In H. Mikkelson & R. Jourdenais (Eds.), *The Routledge Handbook of Interpreting*. New York: Routledge, 368–383.

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics*, 9, 1460-1474.

Kidawara, Y. (2020). Leading the world with speech translation technology. *NICT News Vol. 480*, 1–2.

Matsushita, K. (2020). The results of an online survey on remote interpreting under COVID-19. *Interpreting and Translation Studies, 20*, 125–146.

Moser-Mercer, B. (1996). Quality in interpreting: Some methodological issues. *The Interpreters' Newsletter,* 7, 43–55.

Zwischenberger, C. (2010). Quality criteria in simultaneous interpreting: An international vs. a national view. *The Interpreters' Newsletter 15*, 127–142.

••••

Charlene Meyers
Université de Mons-Hainaut, Belgium
charlene.meyers@umons.ac.be

Translation Strategies Used by Students to Sight Translate Metaphors: A Study Based on the MIPVU and Schäffner's Cognitive Approach

In recent years, pedagogical studies have turned to translation classes to analyze how translation students cope with metaphors (as a pragmatic, linguistic, and cognitive item according to the metaphor theory by Lakoff and Johnson (1980)) in a source text (ST). The pedagogical scope of translation studies has mainly focused on "metaphor translation competence," which is defined as the ability of translators to transfer the mapping of metaphors from the ST into metaphors in a target text (TT) (Sjørup, 2013). Moreover, some studies such as the one by Jensen (2005), have taken an interest in comparing the students' level of metaphor translation proficiency with that of young professionals and experts.

However, few studies provide details about the procedure used to identify metaphors in the ST and the TTs and about the approach or model used to describe the translation strategies employed to translate metaphors. Moreover, whereas Schäffner and Shuttleworth (2013) stress that "there has been more research on metaphor in translation than in interpreting" (p. 13), it is also clear that the study of metaphors in sight translation has been almost completely ignored.

The aim of the present study is to examine the metaphor translation competence of Master students in order to evaluate their translation solutions for metaphors in a ST. This study relies on the MIPVU (Metaphor Identification Procedure Vrije Universiteit Amsterdam developed by Steen et al. in 2010) for metaphor identification and on Schäffner's (2004) cognitive approach to metaphor translation in order to describe the translation strategies used by the students. In addition, the study is also innovative because it is based on sight translation, which is a hybrid between translation and interpreting (Dragsted & Hansen, 2009), instead of written translation.

Metaphor translation competence was examined to provide an answer to the following research questions:

- Are metaphors identified in the ST retained in the TTs?
- Is there a difference in the proportion of metaphors among the TTs?
- Which translation strategies are used according to the TTs? Is the frequency of strategies identified in the TTs equivalent?
- Which translation strategies are used according to metaphors in the ST? Is the frequency of strategies applied to the metaphors in the ST equivalent?

To test metaphor translation competence, the translation of metaphors in a ST by 20 translation students was analyzed. A 495-word-long ST was selected in the *New Scientist* and was examined for metaphors by following the MIPVU (Steen et al., 2010). The metaphor-related words (MRWs) in the ST were coded by two researchers using the identification procedure on every lexical unit of the ST. Intercoder agreement was calculated on the identification of MRWs and showed, after three rounds, a Cohen's kappa of 0.958, which reveals an "almost perfect agreement" (McHugh, 2012, p. 279). In total, 75 MRWs were identified in the ST by both coders and were retained for the experiment.

20 Master students were then asked to sight translate this ST from English to French, as a regular sight translation exercise during their scientific and technical translation class. The recording of their sight translations was transcribed and the MIPVU (Steen et al., 2010) was applied by the same two researchers on the TTs to identify the MRWs. Intercoder agreement was also calculated on the TTs and showed "substantial agreement" ($0.61 < \kappa < 0.80$) for some TTs and "almost perfect agreement" ($0.81 < \kappa < 1.00$) for others.

Based on Schäffner's (2004) cognitive approach to metaphor translation, 9 types of translation strategies were identified (3 of which were created for the coding experiment because absent of the original approach) and coded for the translation of each MRW in each TT (total = 1500 tokens): such as unsuccessful transfer (of MRW) at the micro-level, conceptual equivalence between the ST and the TT but utilization of a different aspect of metaphor, non-metaphorical translation strategy, perfect match between the MRW in the ST and the MRW in the TT, etc.

The results show that the translation of MRWs in the ST is metaphorical in 80.26% of cases in the TTs and that there is no significant difference in the proportion of MRWs between TTs (c^2 (19) = 25.186, $p = 0.1545$), which shows that students tend to deal with metaphors in a similar way. Moreover, the

coding of the translation strategies shows that the most used strategy is the perfect match (53.13%) and that there is a significant difference in the proportion of strategies used between the TTs (c^2 = 222.84, p = 0.0004998. Cramer's V = 0.146). Most interestingly, there is a significant difference in the proportion of translation strategies applied to individual MRWs (c^2 = 2744.8, p = 0.0004998. Cramer's V = 0.511), which indicates that some strategies were overused and others underused in the translation of certain MRWs.

References

Dragsted, B., & Hansen, I. G. (2009). Exploring Translation and Interpreting Hybrids. The Case of Sight Translation. Meta, 54 (3), 588–604.

Jensen, A. (2005). Coping with Metaphor: A cognitive approach to translating metaphor. *HERMES - Journal of Language and Communication in Business, 18* (35), 183–209.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22* (3), 276–282.

Schäffner, C. (2004). Metaphor and translation: Some implications of a cognitive approach. Journal of Pragmatics, 36 (7), 1253–1269.

Schäffner, C., & Shuttleworth, M. (2013). Metaphor in translation: Possibilities for process research. Target, 25 (1), 93–106.

Sjørup, A. C. (2013). *Cognitive effort in metaphor translation: An eye-tracking and key-logging study* [Doctoral dissertation]. Copenhagen Business School.

Steen, G., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T., & Trijntje, P. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. Amsterdam: John Benjamins.

••••

Zoë Miljanović
RWTH Aachen University
zoe.miljanovic@ifaar.rwth-aachen.de

A Corpus for the Investigation of Literary Translation Over Time

Corpus-based studies of literary translation often focus on individual authors or works, specific features and questions of style (Bosseaux 2007; Winters 2009; Saldanha 2005) or translation properties (Hansen-Schirra 2011; Ben-Ari 1998; Øverås 1998). But there is still room to investigate literary translation more broadly over time regarding the involved language and literary systems and language change (cf. Xia 2014). Among the multiple constraints involved in translation, such as the original, the involved languages and audiences or socio-cultural backgrounds, the relation between target language change and the translated language is also highly relevant (Ji 2012). Translation may even influence the target language (Johansson 2007: 33). The relation between translation and language change could be investigated with a diachronic corpus that offers larger quantitative evidence and may also reveal the linguistic consequences of a changing concept of translation over time as well as contribute to theory-building (cf. Hansen-Schirra 2011: 143; Apel 1983: 31-32). This paper presents a corpus for the language pair English-German and the methodology employed for such a project.

The study is grounded in descriptive, corpus-based translation studies. It draws on polysystem theory, according to which the default position of translation is peripheral: It conforms to the norms of the target system without exerting any influence (Even-Zohar 1990; Kruger 2012). Whether translation is in fact a 'factor of conservatism' (Even-Zohar 1990: 48) relates to what Holmes (1988) calls the cross-temporal factor: Questions of this diachronic study of literary translation are whether translation leads or follows change, and whether the translator modernizes or historicizes the language (Hermans 2020 [1999]: 26; cf. Popovič's typology of cultural time (Špirk 2009)). How does the temporal distance between original and translation influence linguistic characteristics? Are newer translations of older texts more similar to translations of contemporary literature or the contemporary target language norms despite the language of the original?

To investigate the relation between target language change and translated language and its properties in comparison to the source, a diachronic corpus is built. The corpus is bi-directional and parallel. This allows not only the comparison of the source texts with their translations but also the monolingual comparison between originals and translations and the multilingual comparison of originals and translations between languages (cf. Johansson 2003; Neumann & Hansen-Schirra 2012). This corpus design makes it easier to assess whether findings are translational features or due to contrastive differences, which already result in differences and constrain the translators' choices. The corpus is specialised and includes only narrative fiction of various genres. It represents an original and translated part of the English and German literary systems. The originals were published between 1750 and 1950 and translated within that same timeframe. This period has until recently been relatively neglected with regard to language change. In both languages, it is characterised by ongoing standardisation processes (Tieken-Boon van Ostade 2009; von Polenz 1999). If the changing lexico-grammatical repertoire is not established in the target system, translated language may be more conservative. Additionally, the linguistic effects of changing translation norms and practices over time may be empirically observed with such a corpus (cf. Bassnett 2002; Bastin & Bandia 2006; Albrecht & Plack 2018).

Presently, the collection contains 100 full original novels, novellas and stories from 50 authors with their translations in each translation direction. The majority of originals and translations spreads over the second half of the 200 years period, but the corpus is not yet complete. It will be divided into periods of 20 years to investigate the development of the language and features of literary translation. Various factors may influence the language of the translations, as there are different, possibly competing 'pulls' at play (Halverson 2017). In a diachronic move, the source texts enter the target literary system as translations, which can also draw on features from originals in the target system. At the same time, the translated texts may have a function in and influence target subsystems (see figure 1; cf. Santoyo 2006; Levý 2011).

Apart from the corpus methodology, this paper presents a first characterisation of the linguistic features of the corpus, its subcorpora and chronological divisions in terms of shallow features, namely part-of-speech distributions, lexical density, standardised type-token ratio, mean sentence- and word-length (cf. Holtz 2011: 66; Steiner 2012), in order to capture contrastive

differences between the originals and compare the translations. For a higher comparability of the part-of-speech distributions, universal part-of-speech tags are used (De Marneffe et al. 2021). These generalisations reveal possible translation features, which continue to be of interest in corpus-based translation studies (Laviosa & Liu 2021), such as tendencies to conform to target language norms or to retain source text features (cf. Halverson 2017; Hansen-Schirra 2011; Toury 1995), and may differ depending on the translation direction. Any patterns will be analysed more closely in the larger project.
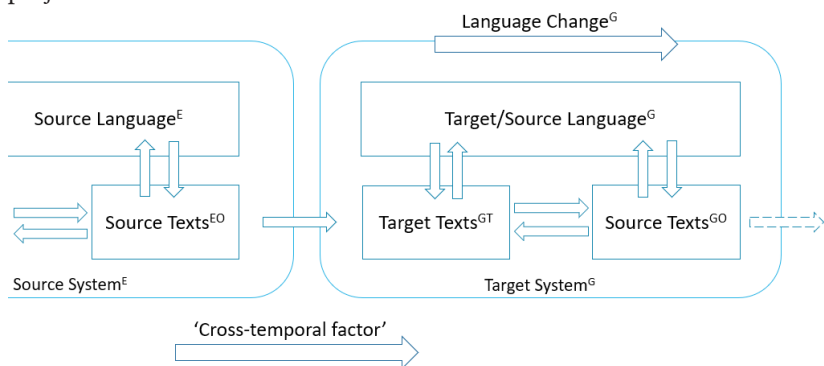
Figure 1. Conceptual model of a snapshot of source and target system in one translation direction

References

Albrecht, J., and I. Plack (2018). *Europäische Übersetzungsgeschichte*. Tübingen: Narr.

Apel, F. (1983). *Literarische Übersetzung*. Stuttgart: Metzler.

Bassnett, S. (2002). *Translation Studies*. London, New York: Routledge.

Bastin, G. L., and P. F. Bandia (Eds.) (2006). *Charting the Future of Translation History*. Ottawa: University of Ottawa Press.

Ben-Ari, N. (1998). The ambivalent case of repetitions in literary translation. Avoiding repetitions: a "universal" of translation? *Meta,* 43 (1), 68-78.

Bosseaux, C. (2007). *How Does It Feel? Point of View in Translation: The Case of Virginia Woolf into French*. Leiden: Brill.

De Marneffe, M.-C., C. D. Manning, J. Nivre, and D. Zeman (2021). Universal dependencies. *Computational Linguistics*, 47 (2), 255-308.

Even-Zohar, I. (1990). The position of translated literature within the literary polysystem. *Poetics Today*, 11 (1), 45-51.

Halverson, S. (2017). Gravitational pull in translation: testing a revised model. In G. de Sutter, M.-A. Lefer, & I. Delaere (Eds.), *Empirical Translation Studies*. Berlin: De Gruyter, 9-45.

Hansen-Schirra, S. (2011). Between normalization and shining-through: Specific properties of English-German translations and their influence on the target language. In S. Kranich, V. Becher, S. Höder, & J. House (Eds.), *Multilingual Discourse Properties*. Amsterdam: John Benjamins, 135-162.

Hermans, T. (2020 [1999]). *Translation in Systems. Descriptive and Systemic Approaches Explained.* London, New York: Routledge.

Holmes, J. (1988). *Translated! Papers on Literary Translation and Translation Studies*, R. van den Broeck (Ed.). Amsterdam: Rodopi.

Holtz, M. (2011). Lexico-grammatical Properties of Abstracts and Research Articles. A Corpus-based Study of Scientific Discourse from Multiple Disciplines. PhD Thesis, Technische Universität Darmstadt.

Ji, M. (2012). Hypothesis testing in corpus-based literary translation studies. In M. P. Oakes & M. Ji (Eds.), *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam: John Benjamins, 53-72.

Johansson, S. (2003). Contrastive linguistics and corpora. In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi, 31-44.

Johansson, S. (2007). *Seeing Through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.

Kruger, H. (2012). *Postcolonial Polysystems. The Production and Reception of Translated Children's Literature in South Africa*. Amsterdam: John Benjamins.

Laviosa, S., and K. Liu (2021). The pervasiveness of corpora in translation studies. *Translation Quarterly,* 101, 5-20.

Levý, J. (2011). *The Art of Translation*. Amsterdam: John Benjamins.

Neumann, S., and S. Hansen-Schirra (2012). Corpus methodology and design. In S. Hansen-Schirra, S. Neumann, & E. Steiner (Eds.), *Cross-Linguistic Corpora for the Study of Translations*. Berlin: De Gruyter, 21-33.

Øverås, L. (1998). In search of the third code: An investigation of norms in literary translation. *Meta,* 43 (4), 557-570.

Saldanha, G. (2005). Style of Translation: An Exploration of Stylistic Patterns in the Translations of Margaret Jull Costa and Peter Bush. PhD Thesis, Dublin City University.

Santoyo, J.-C. (2006). Blank spaces in the history of translation. In G. L. Bastin, & P. F. Bandia (Eds.), *Charting the Future of Translation History*. Ottawa: University of Ottawa Press, 11-43.

Špirk, A. (2009). Anton Popovič's contribution to translation studies. *Target*, 21 (1), 3-29.

Steiner, E. (2012). A characterization of the resource based on shallow statistics. In S. Hansen-Schirra, S. Neumann, & E. Steiner (Eds.), *Cross-Linguistic Corpora for the Study of Translations*. Berlin: De Gruyter, 71-89.

Tieken-Boon van Ostade, I. (2009). *An Introduction to Late Modern English*. Edinburgh: Edinburgh University Press.

Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.

Von Polenz, P. (1999). *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart. Band III: 19. und 20. Jahrhundert*. Berlin: De Gruyter.

Winters, M. (2009). Modal particles explained: How modal particles creep into translations and reveal translators' styles. *Target*, 21 (1), 74-97.

Xia, Y. (2014). *Normalization in Translation: Corpus-based Diachronic Research into Twentieth-Century English-Chinese Fictional Translation*. Newcastle: Cambridge Scholars Press.

••••

Jean Nitzke
University of Agder, Norway
jeann@uia.no

Carmen Canfora
University of Mainz, Germany
canfora@uni-mainz.de

Silvia Hansen-Schirra
University of Mainz, Germany
hansenss@uni-mainz.de

Decisions in post-editing projects: Using semi-structured interviews with stakeholders from the language industry to update a decision tree model for post-editing tasks

Post-editing (PE) machine translation has become increasingly important in the professional translation industry in recent years. However, not every translation job is suitable for machine translation (MT) and different aspects have to be taken into account to decide for or against PE. The continuum of options for translation/post-editing projects is broad: no PE, light PE, full PE, full PE plus revision or human translation, i.e., translation from scratch with CAT tools, with or without further quality assurance measures like revision or automatic quality assurance. The variety of factors that potentially influence the outline of a project was recently published as a decision tree for post-editing jobs that shall guide decision makers in deciding whether a job is suitable for post-editing and if so what kind of quality assurance might lead to fit-for-purpose translations (Nitzke et al. 2019). The different branches of the decision tree cover aspects such as the availability of an MT system, characteristics of the source text (type), specificities of the target culture, time pressure, budget, etc.

This approach was based on a risk and process analyses (for a risk management framework, see ISO 31000:2018), which allows predictions on the potential but also on the risks of using MT and PE (Canfora and Ottmann 2020). However, the rather theoretical assumptions underlying the model present an idealised approach to post-editing projects and the market reality

may diverge. Therefore, our current research aims at comparing empirical evidence from market stakeholders with our existing model and assess the potentials to feed our model with data from realistic processes and environments, even though processes established in every-day practise may not be prototypical.

Hence, we developed a semi-structured interview building on our theoretical approach addressing language service providers, project managers, and institutions that work with post-editing projects and make the decisions reflected in our model. The interview consists of 20 open questions and a scoring task, which will allow us to combine both qualitative and quantitative analyses. The questions cover aspects like the role of text types, risk awareness and risk assessment, target recipients, final quality requirements, money and time frame and more aspects conceptualising the post-editing project. We also ask about the interviewees attitude towards and the availability of MT systems, if they use MT within a CAT tool, as well as their best practise concerning guidelines.

Based on 19 interviews, we will describe and rank (descriptive statistics and qualitative analysis) which of the assumed aspects that influence the PE decision making process are indeed important in real-life projects, how they intertwine, and how risk management considerations as well as liability issues affect these decisions. The findings provide innovative insights on how the translation service industry deals with PE projects, whether there are similar patterns between different actors, and, finally, how to adjust our model. In this scope, we have to discuss potential mismatches between an idealised model and the reality established at the translation market, e.g., the differences between what role the budget for a translation/post-editing project should play ideally and does play in real projects.

In the presentation, we will discuss the existing decision tree model, the interview scheme, and the analysis pattern we developed for this study. We will present anonymised information on our participants taken from the meta data collected within our study and findings concerning their general MT use, their post-editing processes, their risk awareness, the potential market pressures concerning time and price constraints and their generalised decision-making procedures. Finally, we will feed back these results into the existing model and present the updated decision tree for post-editing tasks. Further, we will be discussing implications for curricula contents both for post-editing and project management competences.

References

Canfora, C., and A. Ottmann. (2020). "Risks in neural machine translation." *Translation Spaces* 9(1), 58-77.

ISO (International Standardization Organization). (2018). *ISO 31000:2018 Risk management – Guidelines.* Geneva: ISO.

Nitzke, J., S. Hansen-Schirra, and C. Canfora. (2019). "Risk management and post-editing competence." *The Journal of Specialised Translation* 31, 239-259.

....

Alice Ray
Université d'Orléans, France
alice.ray@univ-orleans.fr

"Get Schwifty": Translating Rick & Morty's Science Fictional World

Since 2013, Dan Harmon and Justin Roiland's postmodernist TV show *Rick and Morty* offers a science fictional world filled with dark and nonsense humour and a tinge of nihilistic philosophy to the audience. The success of the animated TV show lies in its originality and its dark portrait of humanity, society, and life; through humour, sarcasm and parody, the show addresses various deep philosophical questions (Abesamis, 2019). However, it is first and foremost a science fiction show and the authors use all the genre's tools, including linguistic creativity (Angenot, 1979; Cheyne, 2008, Delany, 2009) to ground their original world. Indeed, one of the specificities of science fiction is the creation of new words and phraseologies, which are linguistic hybrids: they are created both as artistic tools (creating a universe of its own) and as scientific terminology, with plausible terms that *could be* integrated in our lexicon *if* the world were as the authors paint it (*conapt, servok, biot*, etc.). Fiction terms are used in three different ways throughout the show: first, as a way to create and ground the science fictional world. Even though *Rick and Morty* is rooted in parody, for the humour of the show to function, the audience needs to believe in the created universe and identify it as being science fictional. Then, fiction terms are also used as a parody of themselves and science fiction genre in general. And finally, science fiction has always been an interreferential genre: not only novums and fiction terms create a megatext (Broderick, 1992), but sci-fi works tend to be generically intertextual (Bréan, 2019). In *Rick and Morty*, fiction terms are also used as a way to parody other pop culture works (*Jurassic Park, The Matrix,* Disneyland, etc.). Thus, invented terms and phraseologies are an important humoristic tool in the show. They allow the authors to build a stable world while offering another way of including parodic humour in the scripts. However, although comedy can be a true challenge for translators as it is deeply rooted in culture and language, science fictional comedy is a double challenge as it is both culturally rooted and cognitively estranged. Moreover, in the case of

audiovisual translation, translators need to transfer the humoristic paradigm of the fiction terms while complying with dubbing constraints of time and labial synchronisation (even though it is less obvious in animated fiction). According to the typology of *Rick and Morty*'s fiction terms previously established, this submission offers a qualitative and quantitative study of the French translation of the show's fiction terms. All the terms will be extracted from *Rick and Morty*'s seasons 1 to 4 (41 episodes) in their original version and their French dubbed translation. They will be analysed through contrastive linguistics (lexical and phraseological patterns of fiction terms in English and in French) and translation studies approaches in order to understand the complexity that arises from translating at once referentiality, lexical creativity and humour for an audiovisual media. Moreover, if humour and dubbing have already been studied, little has been said about science fiction humour and its translation (although some works study humour in science fiction [Candelaria, 2018]). This presentation also aims to identify the specificities of science fiction humour and the way it was transferred into French, by using linguistic criteria to categorize fiction terms' translation (Ray, 2019) and to classify some translation patterns in science fiction humour. The preliminary results of the study show a subdivision of fiction terms' translation in three groups: 1. equivalent effect on the target audience (whether it is humoristic or to ground the science fictional world). Through the analysis, the notion of "equivalent effect" (or "functional equivalence" in Nida's terms [1964]) will be questioned as a reliable notion for fiction terms (especially regarding the elimination of everything related to the source culture as *Rick & Morty* is specifically grounded in North American culture). 2. Change or loss of the effect on the target audience (whether lost in translation or modified because of dubbing constraints, cultural references, or lexical choices). 3. Amplification of the effect on the target audience (in some rare cases, the translation amplifies the humoristic feature of the fiction terms in the target language). Thanks to contrastive linguistics tools (semantic and morphosyntactic analysis in both languages) and translation studies approach (translation of humour [Gulyás, 2021, Jankowska, 2009, Martínez-Sierra, 2006] and the functionalist theory of translation [Reiss, 2009, (1995)]), the study aims to a better understanding of both fiction terms and TV comedy translation as the lexical creativity of *Rick and Morty* is anchored in its humour. Both theories will be used to determine how the linguistic construction of the fiction terms and their specific function in the text shape the translation choices. In other words: "It's time to get schwifty".

References

Abesamis, L. C. (2019). and Wayne Yuen. *Rick and Morty and Philosophy. In the Beginning was the Squanch.* Chicago: Open Court Publishing.

Angenot, M. (1979). The Absent Paradigm: An Introduction to the Semiotics of Science Fiction. *Science Fiction Studies*, 6 (17).

Asimakoulas, D. (2004). Towards a model of describing humour translation. A case study of the Greek subtitled versions of *Airplane! and Naked Gun. Meta*, 49.

Baker, M. and Saldanha G. (2021). *Routledge Encyclopedia of Translation Studies 3rd edition.* London: Routledge.

Bucaria, C. (2008). Dubbing Dark Humour: A Case Study in Audiovisual Translation. *Lodz Papers in Pragmatics*, 4 (2), 215-240.

Bréan, Simon. (2019). Dissonance et harmonies culturelles: la mise en crise des sociétés dans la science-fiction française contemporaine (2010-2019). *Oeuvres et Critiques* XLIV (2).

Broderick, D. (1992). Reading SF as a Mega-Text. *The New York Review of Science Fiction*, 47, 8-11.

Candelaria, R. L. (2018). Absurdist Sci-Fi Humor: Comparable Attitudes Regarding Absurdism in *Hitchhiker's Guide to the Galaxy and Rick and Morty, OUR Journal: ODU Undergraduate Research Journal*, 5. URL: https://digitalcommons.odu.edu/ourj/vol5/iss1/1/

Cheyne, R. (2008). Created Languages in Science Fiction. *Science Fiction Studies*, 35 (3), 386–403.

Delany, S. R. (2009). *The jewel-hinged jaw: notes on the language of science fiction.* Wesleyan University Press.

Evans, T. (2015). Wubba Lubba Dub Dub!: The Pursuit of Happiness in *Rick and Morty. Under Construction @ Keele,* 2 (1). URL: https://viewer.joomag.com/under-construction-keele-volume-2-issue-1/0077049001458827133?page=18.

Fuller, A. T. (2020). *An Absurd Theory of Humor.* Dickinson College Honors Theses.

Granger, S., and Meunier F. (2008). *Phraseology: An interdisciplinary perspective.* Amsterdam: John Benjamins Publishing Company.

Gulyás, A. (2021). Lexical Creativity and Humor in Translation: On Rabelais' Linguistic Genius and the Difficulties in Translating His Works. *Lexis. Journal in English Lexicology*, 17 (17).

Harmon, D., and Roiland J. (2013) *Rick and Morty.* Adult Swim.

Jankowska, A. (2009). Translating Humor in Dubbing and Subtitling. *Translation journal*, 13 (2). URL: https://translationjournal.net/journal/48humor.htm.

Koltun, K. (2018). Rick, Morty, and Absurdism: The Millennial Allure of Dark Humor. *The Forum: Journal of History*, 10 (1). URL: https://digitalcommons.calpoly.edu/forum/vol10/iss1/12.

Loison-Charles, J., and *al. (2019) Du jeu dans la langue : Traduire le jeu de mots*. Villeneuve d'Ascq: Presses universitaires du Septentrion.

Low, P. A. (2011). Translating jokes and puns. *Perspectives*, 19 (1), 59-70.

Martínez-Sierra, J. J. (2006). Translating Audiovisual Humour. a Case Study. *Perspectives*, 13 (4), 289-296.

Mollica, F. and Ramusino P. C. (2020). *Contrastive Phraseology: Languages and Cultures in Comparison*. Cambridge: Cambridge Scholars Publishing.

Nida, E. A. (1964). *Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating*. Brill Archive.

Ray, A. (2019) *Traduire les termes du futur : Analyse du traitement des termes-fictions dans la traduction de l'anglais au français de la littérature de science-fiction*. Université d'Orléans. URL: https://theses.univ-orleans.fr/public/2019ORLE3003_va.pdf.

Reiss, K. (2009 [1995]). *Problématiques de la traduction*. Traduit par Catherine A. Bocquet. Paris: Editions Economica.

Zabalbeascoa, P. (1996). Translating Jokes for Dubbed Television Situation Comedies. *The Translator*, 2 (2), 235-257.

••••

Rossella Resi
Universität Innsbruck, Austria - Università degli Studi di Verona, Italy
rossella.resi@uibk.ac.at

Investigating translation strategies for prenominal complex adjectives within a German-Italian parallel corpus

This paper aims at investigating a parallel corpus of German and Italian texts in order to examine the behaviour of German complex prenominal adjectives as well as possible translation strategies in the Italian language, which is characterized by a post-nominal attribution.

While the first part of this work regards parallel corpora as suitable sources of data for quantitatively and descriptively investigating a big syntactical difference between the two languages (following the framework of corpus-based contrastive linguistics described in Laviosa, 1998 and 2003; Ebeling & Hasselgard, 2018; Johansson, 1997, 2003, 2011); on the second part, a corpus-driven translative analysis is conducted to spot statistically significant translation strategies and test the quality of exceptional translation behaviours as well as their possible generalisation (following the framework of DTS: Baker 1993, Toury, 1991a, 1991b; Tognini-Bonelli, 2000 and 2001; Laviosa, 2002)

The corpus was compiled as a parallel corpus on Sketch Engine (Kilgariff et al. 2004) allowing to navigate each language also separately. A first examination of the corpus of German texts brought to a detailed subcategorization of German prenominal complex adjectives. All prenominal structures which do not find an equivalent syntactic structure in the Italian language were extrapolated and systematized. It was immediately clear how prolific and diversified this kind of German structures is, which includes predicative structures with past participle or present participle; compound adjectives; more than one non-coordinated prenominal adjectives; adjectives modified by adverbs or by more complex structures; graded adjectives; infinitive structure.

A statistically significant equivalence of these structures to be found in the Italian texts is the position of the attribution after the nominal head: in the form of a relative clause, a post-nominal adjective or a postnominal adjectival construction.

In the following phase, we take a closer and qualitatively look at a statistically insignificant translation strategy called transposition (Alcaraz & Hughes, 2002; Kather, 2011), where the grammatical category of a syntactic element is changed and dependencies amongst syntactical elements are restructured.

    a. *die [im Anhang II **festgelegten**] Kategorien*
       *la **definizione** delle categorie di cui all'allegato II*

We suppose that since the complexity of the structure between determiner and substantive requires a syntactical reference in both directions, the semantic information within the complex structure is in some cases elevated to a higher semantic level within the sentence which must be taken into account when translating it into Italian. Provided that the context allows this usage, we believe that transposition is a fully plausible and yet not enough exploited strategy for the translation of prenominal complex adjectives.

We investigate therefore the possibility of using transposition strategy also for other cases rather than those already found in the corpus, nominalizing the semantic important information included in the prenominal adjectival structure in order to account for the semantic shift.

    b. *…, dass die [auf der Grundlage der Marketingverträge **erworbenen**] Dienstleistungen…*
       *che **l'acquisizione** di servizi tramite contratti di marketing*

In the talk I will present examples where transposition can be used to translate complex adjectives and cases in which transposition is even to prefer. We will also try to see if there are syntactical circumstances which favour the use of transposition.

References

Alcaraz V. E. & B. Hughes. 2002. *Legal Translation Explained.* Manchester: St Jerome Publishing.

Baker, M. 1993. "Corpus Linguistics and Translation Studies – Implications and Applications." *Text and Technology in Honour of John Sinclair* ed. by M. Baker, G. Francis and E. Tognini-Bonelli. 233-250 Philadelphia/Amsterdam: John Benjamin Publishing Company.

Ebeling, S.O. & H. Hasselgard. 2018 "Corpora et comparatio linguarum: Textual and contextual perspectives" *Bergen Language and Linguistics Studies 9 (1).* https://doi.org/10.15845/bells.v9i1.1518.

Johansson, S. 1997. „Using the English-Norwegian Parallel Corpus – a corpus for contrastive analysis and translation studies" *PALC ‚97 Practical Applications in Language Corpor*a ed. By B. Lewandowska-Tomaszczyk and P. J. Melia. LÆdz: LÆdz University Press. 282-296.

Johansson, S. 2003. "Corpora and Translation Studies." *Corpus -based Approaches to Contrastive Linguistics and Translation Studies.* Ed. By S. Granger, J. Lerot and S. Petch-Tyson. Amsterdam: Rodopi. 45-54

Johansson, S. 2011. "A multilingual outlook of corpora studies." *Perspectives on Corpus Linguistics* ed. by V. Viana, S . Zyngier and G. Barnbrook, 115–129. Amsterdam: John Benjamins.

Kather, L.C. 2011. *Mediazione Linguistica Tedesco-Italiano. Aspetti teorici e applicative. Esempi di strategie traduttive. Casi di testi tradotti.* Heopli: Milano

Kilgarriff, A. et al. 2004. "ITRI-04-08 the sketch engine." *Information Technology.* 105-116.

Laviosa, S. 1998. "The Corpus-based Approach: A New paradigma in Translation Studies." *Meta Journal des traducteurs/ Translators Journal*, vol. 43, n. 4. 474-479.

Laviosa, S. 2002 "Corpus-based Translation Studies. Theory, Findings, Applications." *Approaches to Translation Studies. Vol.17* ed. by R. van den Broeck & T. Naaijkens. Amsterdam: Rodopi.

Laviosa, S. 2003. "Contrastive Linguistics and corpora." *Corpus -based Approaches to Contrastive Linguistics and Translation Studies.* Ed. By S. Granger, J. Lerot and S. Petch-Tyson. Amsterdam: Rodopi. 31- 44

Tognini-Bonelli, E. 2000. "Unità funzionali complete" in inglese e in italiano: verso un approccio "corpus-driven." *Corpora nella Didattica della Traduzione.* Ed. by S. Bernardini and F. Zanettin. Bologna: CLUEB. 153-174.

Tognini-Bonelli, E. 2001. "Corpus Linguistics at work." *Studies in Corpus Linguistics. 6.* 224. Amsterdam: John Benjamins Publishing Company

Toury, G. 1991a. "What are Descriptive Studies into Translation Likely to Yield apart from Isolated Descriptions." *Translation Studies: The State of The Art* ed. by K. M. van Leuven-Ywart & T. Naaijkens 179-92. Amsterdam: Rodopi.

Toury, G. 1991b. "Experimental in Translation Studies: Achievements, Prospects and Some Pitfalls". *Empirical Research in Translation and Intercultural Studies* ed. by Sonja Tirkkonen-Condit, 45-66. Tübingen: Gunter Narr.

....

Matt Riemland
Dublin City University
matthew.riemland2@mail.dcu.ie

Language Status and Loanwords in Translation: Determining the correlation between source-language status and loanwords in translated texts using corpora

This presentation's research question asks: does higher source-language status correlate with a higher number of loanwords in translation? The study adopts Edwards' (1996, 703) sociolinguistic framing of language status, which holds that "the status of a language is its position vis-à-vis others". This relative positioning may also be characterized as a hierarchy of language power. Translation scholars have often considered the ways in which discrepancies between the status or power of source languages and target languages may influence translated texts and the translation process. Marais (2014, 187) notes that the discipline "has focused, similar to literary studies, postcolonial studies, and even history, on power struggles".

Some scholars have asserted a correlation between source-language status and source-language influence on translated texts, intending for corpus-based research to test this claim. Baker (1993, 183) suggests that translations from relatively higher-status languages may tend to exhibit features less typical of the target language and context. Similarly, Toury (2012, 314) asserts that translations from comparatively high-status languages into comparatively low-status languages tend to exhibit more "interference". That is, the higher prestige of the source or donating language results in translations containing more features typical of the source language. Loanwords or lexical borrowings taken from the source text and reproduced in the target text represent clear evidence of interference. Therefore, this study adopts as its hypothesis the expectation that translations from comparatively high-status languages will contain more loanwords than translations from comparatively low-status languages.

To test this hypothesis, this study utilizes comparable corpus methodology, constructing a multilingual corpus consisting of approximately 26 million words. The corpus texts are literary prose, and are primarily historical

works (late 19th century to early 20th century) in the public domain. The selected languages include English and French as high-status languages, German and Italian as medium-status languages, and Croatian, Swedish, and Irish as low-status languages. Each language has its own comparable subcorpus of original texts. There is a subcorpus for translations into each language from the other selected languages. The multilingual corpus contains numerous translations for each possible status discrepancy between source language and target language, e.g. high-status language to low-status language translations.

This study identifies what may be termed "translational loanwords". It defines translational loanwords as word forms in the translated texts that are reasonably suspected to be borrowed from the texts' respective source languages. Loanwords originating from languages other than the texts' source languages – e.g. French loanwords in German source texts that are reproduced in English target texts – are not counted, as they do not represent direct translational interference in the closed source-target system. The study identifies loanword candidates by comparing the relative frequencies of words in translated corpora for a given language pair with the relative frequencies of these same words in the comparable corpora constructed for the language pair's source and target languages. Concordances for each possible loanword candidate are then manually reviewed in order to distinguish true loanwords from noise such as proper nouns. Finally, loanword totals for each possible status discrepancy between source language and target language are compared to determine whether higher source-language status correlates with a higher number of loanwords in translation.

This study applies quantitative methodology to a topic that previous work has approached qualitatively or theoretically. It empirically tests for a causal relationship that has been asserted by Baker and Toury, two of the field's most influential scholars. This presentation is intended to serve as a model for further research on the potential correlation between language status and measurable evidence of interference – including but not limited to loanword usage – in translation.

References

Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In G. Francis, E. Tognini-Bonelli, and M. Baker (Eds.), *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins Publishing Company, 233–50.

Edwards, J. (1996). Language, prestige and stigma. *Kontaktlinguistik*, 703-708. https://www.degruyter.com/document/doi/10.1515/9783110132649.1.6.703/html.

Marais, K. (2014). Translation Theory and Development Studies: A Complexity Theory Approach. New York: Routledge.

Toury, G. (2012). *Descriptive Translation Studies - and Beyond: Revised Edition.* 2nd ed. Amsterdam/Philadelphia: John Benjamins Publishing Company.

••••

María Isabel Rivas Ginel
Université de Bourgogne Franche-Comté & Universidad de Valladolid
isabelrivasginel@gmail.com

Sarah Theroine
Université de Bourgogne Franche-Comté
Sarah.Theroine@u-bourgogne.fr

Éric Poirier
Université du Québec à Trois-Rivières, Canada
Eric.Poirier@uqtr.ca

Loïc Barrault
Le Mans Université
loic.barrault@univ-lemans.fr

Felix Hoberg
Leipzig University
felix.hoberg@uni-leipzig.de

Oliver Czulo
Leipzig University
oliver.czulo@uni-leipzig.de

Eva Vanmassenhove
Tilburg university
Vanmassenhove@tilburguniversity.edu

Izabella Thomas
Université de Bourgogne Franche-Comté
izabella.thomas@univ-fcomte.fr

Lucie Bernard
Université de Bourgogne Franche-Comté
Lucie.Bernard@u-bourgogne.fr

Playing with gender. All-inclusive games machine translation: The All-inGMT project

The video game industry, known for its constant evolution in terms of technology and devices, has recently seen the appearance of non-binary, transgender, and non-gendered characters in AAA games. These changes, along

with the modernisation of players' profiles, reflect the realities of today's society and pave the road ahead to inclusivity. However, the industry's drive for new technologies and its adoption of Neural Machine Translation (NMT) systems result in new hurdles in the race for equality due to issues related to gender bias and a general lack of large datasets that include these new linguistic phenomena (e.g. gender-neutral pronouns). Previous research in the field of video game localisation (Rivas Ginel & Theroine, 2022) has shown that Google Translate and DeepL can cause up to 3,03% and 2,38% of gender-related errors respectively when translating from English, a gender-neutral language, into French. Furthermore, the 2020 European Language Industry Survey's report indicated that "66% of agencies and 44% of in house translation teams expect to invest in it [MT]" (LIND, 2020, p. 50). Video game developers have profited as well from this trend and nowadays Unity, the most widely used game engine in the field (Rivas Ginel & Theroine, 2022), proposes manifold modules that include MT options. In parallel, video game localisation, "a double-blind process (no audiovisual context, no text linearity)" (Bernal-Merino, 2013, p. 117), does not usually provide sufficient information in order to identify the gender of the speaker or the addressee.

All-inGMT, under the project ISITE BFC and the programme «Investissement d'Avenir», aims at creating an NMT system specially trained to favour or integrate neutral solutions or techniques used in inclusive language (epicene terms and neutral pronouns) used for non-binary characters both in English and in French. This one-year-long project is based on an interdisciplinary approach that combines corpus linguistics and artificial intelligence using parallel corpora extracted from games that include non-binary, non-gendered and transgender characters. Our paper aims to analyse and formalise the different rules and translation techniques in order to apply them in the creation of texts that offer neutral and inclusive solutions. Therefore, it will provide a general comparison of said rules and techniques and how they apply in each working language and showcase examples of constructions and linguistic phenomena that favour neutral and inclusive solutions in their context. Furthermore, the games — translated into French by human translators using neutralisation methods — constitute a conceptual database that represents the video game field and a linguistic data set of neutralisation techniques for this language pair. Additionally, before proceeding to the compilation, we annotated and POS tagged them manually using Taguette by feeding additional information specifically concerning non-binary or

non-sexualized characters (Saunders, Sallis & Byrne, 2020). The annotation phase allowed us to gain more granularity in our analyses and to pre-establish some rules to teach the NMT. These procedures steam from the results of previous research aiming to mitigate gender bias in NMT training by including specific tags to convey gender at a syntactic level (Vanmassenhove et al., 2018) and automatic rewriting in order to create neutral alternatives (Vanmassenhove et al., 2021).

By summer, the project will reach the evaluation phase of the machine's output using metrics such as COMET, ChrF, BLEU or METEOR by adapting and modifying their criteria to fit our purposes. The first results about its performance in terms of domain specialisation, linguistic accuracy, and the system's compliance with the newly created and implemented neutralisation rules should show a substantial improvement when compared and evaluated against state-of-the-art baseline systems (Google Translate and DeepL) in terms of (i) overall translation quality and, (ii) a more in-depth analysis focusing specifically on gender (e.g. usage of gender-neutral terms, gendered pronouns/nouns, etc.). Ultimately, All-inGMT aims to create the first NMT specifically trained for inclusive language in video game localisation. Furthermore, this system will become a tool that can be adapted, updated, and recalibrated *ad infinitum* in order to provide the base for a more comprehensive and advanced NMT specially focused on gender bias mitigation.

References

Bernal-Merino, M. A. (2013) *The Localisation of Video Games.* London: Imperial College. PhD Theses.

Language industry platform (LIND) (2020) *Report: 2020 Language Industry Survey* available at: https://ec.europa.eu/info/departments/translation/language-industry-platform-lind_en

Rivas Ginel, M. I.; Theroine, S. (2022) *Machine Translation and Gender biases in video game localisation: a corpus-based analysis.* ffhal-03540605

Saunders, D.; Sallis, R.; Byrne, B. (2020) *Neural Machine Translation Doesn't Translate Gender Coreference Right Unless You Make It* arXiv:2010.05332

Vanmassenhove, E.; Emmery, C.; Shterionov, D. (2021) *NeuTral Rewriter: A Rule-Based and Neural Approach to Automatic Rewriting into Gender-Neutral Alternatives* arXiv preprint arXiv:2109.06105

Vanmassenhove, E.; Hardmeier, C.; Way, A. (2018) Getting gender right in neural machine translation. *In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3003–3008, Brussels, Belgium.

••••

Rana Roshdy
Dublin City University, Ireland
rana.roshdy2@mail.dcu.ie

Dorothy Kenny
Dublin City University, Ireland
dorothy.kenny@dcu.ie

Assimilation or Accommodation? Lexical variation in Islamic financial law in English

Translation plays a vital role in producing knowledge and contributing to mechanisms aimed at preserving languages and cultures. This paper examines the complex relationships between translation and identity building, by exploring the ways in which the English-language discourse on Islamic finance resonates with the mechanism of cultural survival known as 'translational accommodation'. In the context of migration, this strategy has evolved to describe the efforts of immigrants living in foreign societies to maintain affinity with their native language and resist the 'translational assimilation' that forces them to surrender completely to the dominant language (Cronin, 2006, p.52). The growth of the global Islamic finance industry has attracted a large body of literature on Islamic finance as a technical topic.

However, from the perspective of linguistics and translation studies, little attention has been paid to the lexicon that makes up this specialized discourse. As a prototype of the 'cultural translation' paradigm, Islamic finance discourse in English metaphorically embodies the migration of an Arabic-origin culture. Thus, the central question in this study asks how are the signature concepts of Islamic finance expressed in English? Do they remain irreducibly 'foreign' symbols or are they being assimilated into the new host systems?

This study, which forms part of a corpus-based doctoral dissertation on Islamic law in English, is based on a specially constructed corpus of nearly six million words taken from diverse genres on Islamic finance. Each genre either has a performative or a non-performative function based on whether it is intended for legal application or for information purposes only. The performative category comprises applied genres which have a binding force, including 'policy-making instruments' (e.g., standards, policy documents)

and 'other instruments' (e.g., contracts, agreements, application forms), while the non-performative category covers purely informative genres (e.g., books, academic articles, grey literature). This design will identify whether culture-specific concepts are expressed differently across diverse contexts or genres.

To identify the symbols of linguistic hybridity in this Islamic finance corpus, the research adopts the methodological foundations of corpus-based lexical variation as proposed by Delaere and De Sutter (2017), who recommend the triangulation of two models: profile-based correspondence analysis and logistic regression. Within corpus linguistics, Speelman et al. (2003) define a profile-based approach as an instance of onomasiological variation, whereby a particular concept can be expressed through different linguistic means; thus, a profile refers to a set of formal alternatives, synonymous variants, linguistic designations or labels that can be used to designate the same concept. The frequency of such linguistic alternatives, whether relative or absolute, is the object of the corpus query.

Using Sketch Engine software, the analysis will start with a bottom-up query to build a profile set. Firstly, through the 'keywords' function, loanwords that feature in the English-language discourse on Islamic finance will be extracted. Secondly, a 'collocations' analysis will be employed to identify loanwords that have an endogenous counterpart lexeme, meaning an English approximate or a synonymous naming alternative. To establish the data sets, the study will focus on five profiles, each of which will consist of a loanword and up to two endogenous variants. Thirdly, a concordance analysis of the selected profiles will explore the different linguistic features used to render sharia-based financial concepts, in order to explore concomitant translation norms based on linguistic frequency in the corpus and to interpret ideological implications in the light of postcolonial translation agendas, which aim to preserve intangible cultural heritage and promote the representation of minoritized groups on the global map. Described as a multi-feature analysis, the profile method allows for rigorous investigation as it considers the probability of lexical variation in expressing the symbols of cultural hybridity and quantifies multiple features for a conceptual category rather than looking at linguistic features in isolation. Finally, logistic regression will be implemented to measure the influence of the explanatory variables, 'genre' and 'legal function' (i.e., performative versus non-performative function), on the choice between hybridity versus endogeneity of lexemes in this specialized

discourse. Under logistic regression, endogeneity acts as a binary response variable, assuming two values (loanword versus endogenous lexeme), while 'genre' and 'legal function' are predictor variables. This regression model involves searching for the variants of each profile in each of the different genres, where the odd ratios will indicate the tendency of each lexical variant to occur in a particular 'genre', or 'legal function'. The efficacy of the logistic regression analysis is due to its ability to determine the simultaneous impact of different predictor variables on a response variable.

Preliminary results indicate that English-language texts on Islamic finance manifest high levels of postcolonial hybridity through linguistic borrowing and glossing, suggesting that the promotion of sharia as intangible cultural heritage is accompanied by an ideologically driven variety of English that can be labelled as 'Islamgish' (by blending 'Islamic' and 'English').

References

Asad, T. (1986). The concept of cultural translation in British social anthropology. In J. Clifford &, G. Marcus (Eds.), *Writing Culture: The Poetics And Politics Of Ethnography*. Berkeley: University of California Press, 141-164.

Cronin, M. (2006). *Translation And Identity*. London and New York: Routledge Taylor & Francis Group.

Speelman, D., Grondelaers, S. & Geeraerts, D. (2003). Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities,* 37 (3), 317–337.

Delaere, I. and De Sutter, G. (2017). Variability of English loanword use in Belgian Dutch translations. Measuring the effect of source language and register. In G. De Sutter, M. Lefer & I. Delaere, (Eds.) *Empirical Translation Studies: New Methodological And Theoretical Traditions.* Berlin: De Gruyter Mouton, 81-112.

••••

Ali Saeedi
Kent State University, United States
Asaeedi@kent.edu

Longhui Zou
Kent State University, United States
lzou4@kent.edu

The Effect of Orthography and Language Orientation on Translation Effort

Several studies have examined various topics that deal with relations between language specificity and translation difficulty (Campbell, 1999; Gile, 2005, 2008; Wang & Zou, 2018). Orthography has been utilized in research on machine translation and translation learning (Zhang et al., 2004; El Kholy & Habash, 2012; Chakravarthi et al., 2021; Krepel et al., 2021). An increased cognitive effort is to be expected during the translation process between language pairs written with different scripts compared to those with similar scripts (Lacruz et al., 2016). Text direction has been a recurrent topic of interest in studies on bilingualism and reading comprehension (Barrett et al., 2002; Boroditsky, 2001; Chatterjee et al.,1999; Spalek & Hammad, 2005). Generally, text direction and language orientation cause a spatial bias where right-oriented and left-oriented readers exhibit direction-corresponding biases (Mashat, 2017). However, the relations between orthography and text direction, and translation difficulty have received little attention from researchers in translation process research. Utilizing parallel corpora to compare different language pairs in regard to the orthography and text direction and their relations to translation difficulty received even less attention. Hence, this research explores differences in effort in the translation process of language pairs of different scripts compared to language pairs of similar scripts. The research also studies the effect of text direction of the source and target texts on effort in translation. Attempts to find special strategies and techniques of translation for language pairs of different scripts and/or text directions can be justified if a dissimilar effort is introduced compared to language pairs of similar scripts and/or text directions. This will, in turn, open the room for researching and introducing pedagogical approaches to translator training that might align with our findings.

In this study, we compare three parallel corpora available at CRITT TPR-DB (Carl et al., 2016): BML12 (English to Spanish), RUC17 (English to Chinese), and AR19 (English to Arabic). These three datasets share the same collection of English source texts (STs), allowing us to control the ST difficulty, script variation, and text direction. We choose the BML12 dataset to look into the translation process of the translators working with a target language that has a similar script and the same text direction as English. The RUC17 dataset serves as a representation of the translation process for translators working with a target language that has a different script but a similar textual direction to English. We utilize the AR19 dataset to investigate the translation process of the translators working with the target language that differs from English in terms of both script and text direction. The translators for these three datasets are all native speakers of their respective target languages and have English as their second language of command.

We operationalize translation effort among these three datasets by possible indicators of translation effort, namely technical effort, temporal effort, and cognitive effort (Krings, 2001). Translators of the language pair with the same script (en-es) show significantly more technical effort than translators of the language pairs with different scripts (en-zh and en-ar) seen in overall more insertions, deletions, and number of edits per segment. The temporal effort for translators of en-es, however, is less than those of en-zh and en-ar as observed by the logarithm of duration per segment. Moreover, translators of en-es show less variance in pause-word ratio (PWR), number of fixations on both the ST and target text (TT), and total reading time on both the ST and TT compared to en-zh and en-ar translators. This might suggest less variance in the cognitive effort for en-es compared to en-zh and en-ar due to the similarity of English and Spanish scripts. As for text direction, translators of the same text direction (en-ar) language pair exert significantly more technical effort than translators of different text direction (en-es and en-zh) language pairs as illustrated by overall more insertions, deletions, and the number of edits per segment. Also, there is an overall tendency of higher temporal effort for translators of en-ar compared to those of en-es and en-zh as shown by the logarithm of duration per segment. Moreover, translators of en-es and en-zh exhibit less variance in PWR, number of fixations on both the ST and TT, and total reading time on both the ST and TT than en-ar translators. These findings can be attributed to more variance in the cognitive effort for en-ar compared to en-es and en-zh due to the dissimilarity of

text direction between Arabic and both Spanish and Chinese. This might indicate that different text directions of en-ar language pair add another layer of challenge to translators. We conclude that translation of language pairs of different scripts and/or text directions involves more variant cognitive effort compared to language pairs of similar scripts and/or text directions. This more variant cognitive effort might contribute to more difficulty in translation.

References

Barrett, A. M., Kim, M., Crucian, G. P., & Heilman, K. M. (2002). Spatial bias: Effects of early reading direction on Korean subjects. *Neuropsychologia, 40*(7), 1003-1012.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive psychology, 43*(1), 1-22.

Campbell, S. (1999). A cognitive approach to source text difficulty in translation. *Target. International Journal of Translation Studies, 11*(1), 33-63.

Carl, M., Bangalore, S., & Schaeffer, M. (2016). New directions in empirical translation process research. *Heidelberg: Springer International Publishing Switzerland. doi, 10*, 978-973.

Chakravarthi, B. R., Rani, P., Arcan, M., & McCrae, J. P. (2021). A survey of orthographic information in machine translation. *SN Computer Science, 2*(4), 1-19.

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia, 37*(4), 395-402.

El Kholy, A., & Habash, N. (2012). Orthographic and morphological processing for English–Arabic statistical machine translation. *Machine Translation, 26*(1), 25-45.

Gile, D. (2005). Directionality in conference interpreting: A cognitive view. *Communication & cognition. Monographies, 38*(1-2), 9-26.

Gile, D. (2008). Issues in research into conference interpreting. In *Übersetzung* (pp. 767-779). De Gruyter Mouton.

Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes* (Vol. 5). Kent State University Press.

Krepel, A., de Bree, E. H., & de Jong, P. F. (2021). Does the availability of orthography support L2 word learning?. *Reading and Writing, 34*(2), 467-496.

Lacruz, I., Carl, M., Yamada, M., & Aizawa, A. (2016). Pause metrics and machine translation utility. *Proceedings of the NLP*, 1213-1216.

Mashat, A. A. (2017). *An Eye Tracking Study to Investigate the Influence of Language and Text Direction on Multimedia.* Old Dominion University.

Spalek, T. M., & Hammad, S. (2005). The left-to-right bias in inhibition of return is due to the direction of reading. *Psychological Science, 16*(1), 15-18.

Wang, B., & Zou, B. (2018). Exploring language specificity as a variable in Chinese-English interpreting. A corpus-based investigation. In *Making way in corpus-based interpreting studies* (pp. 65-82). Springer.

Zhang, M., Li, H., & Su, J. (2004). Direct orthographical mapping for machine transliteration. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics* (pp. 716-722).

••••

Roser Sánchez-Castany
Universitat Jaume I de Castellón, Spain
castan@uji.es

Thematic analysis approach applied to a corpus of Translation undergraduate degree syllabi

In an era marked, among other things, by globalisation, automation, immediacy, the ubiquity of technologies and the omnipresence of English as a lingua franca, the language industry has changed dramatically in recent decades. Now, the language industry in general, and the translation market in particular, demand new profiles of translators with a marked technological character, which have been changing at a dizzying pace in recent years.

According to the Language Industry Survey 2021, translator training centres affirm they feel under the same pressure as the professional translation market on many issues, particularly in terms of technological challenges. In fact, all the stakeholders that took part in the survey (training centres, companies and freelancers) acknowledge that machine translation (MT) and post-editing (PE) are by far the most prominent trends in the sector. Not surprisingly, training in translation technologies is seen as a key component of translators' continuing professional development, alongside specialised domain knowledge (EUATC, 2021). There is no doubt that technologies make the sector so tremendously dynamic that they impose continuous retraining on the professionals who are part of it. In this sense, university education has tried to adapt to market demands, although at a much slower pace. Nowadays, there is still a significant technological gap between the training offered in translation technologies in Spanish university classrooms and the real technological demands of the professional translation market and language industries. It thus seems necessary to develop tighter cooperation with the industry to expose students to real-life scenarios (TAUS, 2020; van der Meer, 2021).

In the latest models of translation competence (Kiraly, 2013; PACTE, 2014; EMT Network, 2017, among others), technologies –both generic and specifically designed for the translation field– play a transversal role and are given a level of importance similar to that of other competencies. However, university undergraduate curricula do not yet seem to reflect this change

of paradigm. As a result, they do not offer students an optimum level of employability or a practical acquisition of a series of technological skills required by the language market in the near future (Pym and Torres-Simón, 2021; O'Brien and Rossetti, 2021).

Our current research work extends and complements the existing literature. Its main objective is to investigate and define the current situation of training in translation technologies in Spanish universities through an empirical study using qualitative and quantitative methods. This study focuses on analysing the presence of technological contents among the Translation and Interpreting undergraduate degree curricula in Spain for the academic year 2019-2020, as published on the respective websites of the universities, taking module syllabi as the main source of data gathering. Despite the geographical scope of the research is Spain, the matters raised and their implications are common in academic programs worldwide. This contribution aims to present in detail the thematic analysis of data to analyse textual corpora. In this sense, the decision was to carry out an initial qualitative study (phase A), in which a thematic analysis of the data was undertaken in order to exploit a module syllabi corpus; and a second quantitative study (phase B), in which the results obtained in the first phase were quantified. In the qualitative phase (A), the aim was to identify, organise and analyse in detail the technology-based content of the selected syllabi using a thematic analysis of the data. The aim was to discover patterns –or themes– within the corpus by carefully reading and re-reading them so that the phenomenon studied could be understood and interpreted. However, phase A can also be considered inductive since the aim of coding the data was not to classify them according to a pre-established framework but based on the patterns found in the corpus (Braun and Clarke, 2006; Saldanha and O'Brien, 2014).

For this purpose, the different sub-stages of the thematic analysis will be explained in detail: the corpus collection and preparation phase will be detailed, including how the emptying of all the documentation available online was made; the main characteristics of the texts composing the sample of data will be described; and the procedure carried out for the thematic analysis of the data will be thoroughly explained, including details regarding the use of Atlas.ti, a qualitative analysis software. Finally, some of the most relevant data obtained on teaching translation technologies will be presented.

References

Braun, V, and V. Clarke. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

Pym, A., and E. Torres-Simón. (2021). Is automation changing the translation profession?. *International Journal of the Sociology of Language*, 2021(270), 39–57. https://doi.org/10.1515/ijsl-2020-0015

EMT Network. (2017). *Competence Framework 2017*.

EUATC. (2021). *European Language Industry Survey*.

Kiraly, D. (2013). Towards a View of Translator Competence as an Emergent Phenomenon: Thinking Outside the Box(es) in Translator Education. In D. Kiraly, S. Hansen-Schirra, and K. Maksymski (Eds.), *New Prospects and Perspectives for Educating Language Mediators*. Narr Verlag, 197-224.

O'Brien, S., and A. Rossetti. (2021). Neural machine translation and evolution of the localisation sector, Implications for training, *Journal of Internationalization and Localization*, 7:1/2, 95-121.

PACTE. (2014). First results of PACTE group's experimental research on translation competence acquisition: the acquisition of declarative knowledge of translation. *MonTI. Monografías de Traducción e Interpretación*.

Saldanha, G., and S. O'Brien. (2014). *Research Methodologies in Translation Studies*. Routledge.

TAUS. (2020). *Reinventing the Translation Industry. Briefing from the TAUS Virtual Conference 2020*. TAUS.

van der Meer, Jaap. (2021). Translation Economics of the 2020s. A Journey into the future of the translation industry in eight episodes. *MultiLingual*. https://multilingual.com/issues/july-august-2021/translation-economics-of-the-2020s/.

••••

KILIAN SEEBER
University of Geneva, Switzerland
kilian.seeber@unige.ch

DONGPENG PAN
University of Geneva, Switzerland
dongpeng.pan@unige.ch

What remote interpreters react to: A psychophysiological field study.

While tightly controlled laboratory experiments remain the gold standard for establishing causal relationships between variables, complex phenomena often elude replication in the white room (Black 1955). This is true for experimental research into interpreting, which can only aspire to come up with a reasonable facsimile of reality, inevitably reducing findings to reproducible scenarios. Conversely, traditional qualitative field research methods are suited to report subjective perceptions rather than objective reactions (see Russo, Bendazzoli and Defrancq, 2018). Against this background, we address the question whether it is possible to apply traditional laboratory methods, such as those quantifying participants' psychophysiological responses to carefully designed stimuli (see Rojo Lopez and Korpal, 2020), to the complex and noisy real-life environment in which professional conference interpreters usually work in an attempt at directly observing them in their natural environment.

In order to address this question, we designed an exploratory study aimed at measuring interpreters' psychophysiological responses to naturally occurring stimuli in a realistic online conference environment. A cohort of five professional conference interpreters assigned to a hub for a series of VRI meetings spanning five days, was equipped with mobile sensors collecting real-time electrodermal (EDA) data for the purpose of assessing their level of cognitive and emotional engagement (Chen et al., 2016). Such a method is well-documented for measuring engagement reflecting participants' emotional arousal (Critchley 2002, Sequeir et al. 2009) as well as mental effort (Dawson et al. 2007). The same interpreters were observed over five days, whereby data were collected during the first 30 minutes, in other words, the first turn of each meeting day.

The original audio-video feed of the conference platform was recorded, as was the interpreters' output. Their electrodermal activity was monitored in real-time using Shimmer3 GSR+ wristband sensors, whereby two electrodes were placed on the index finger and middle finger of participants' non-dominant hand. EDA data were recorded at a frequency of 128 Hz. In a two-pass analysis, the audio-video recording of the recording was then screened for salient events based on an inductive approach aimed at finding patterns.

Using an inductive approach, eight salient events were identified in the conference recording, comprising two non-duration type events (starts and shifts) as well as six duration type events (pauses, noise, unclear sound, coughing, text and video).

The EDA signal collected during these events was then down-sampled to 16Hz and analyzed for co-occurring skin conductance responses (SCRs) in a continuous decompositional analysis (CDA, Benedek & Kaernbach 2010) by the extraction of continuous phasic/tonic activity based on standard deconvolution. For non-duration events (meeting start and speaker shift), we selected a three-second time interval of one to four seconds after event onset as the critical time window for SCR analysis. A two-second interval prior to the event onset was used for comparing with the critical window. For duration events, the critical window starts one second after the event onset and ends with the event. We also identified a pre-event window (prior to the onset of the event) and a post-event window (immediately after the offset of the event) half long as the critical window (sprinkled on some occasions to avoid window overlap) to compare with the critical period. The number of significant SCRs within the response window and their amplitude were used as an indicator of engagement. EDA responses during non-duration events reveal that the number of significant skin conductance peaks increases with the beginning of the meeting and with each speaker shift. This could reflect the increased attention of the interpreters facing such changes in a conference.

EDA responses during duration events, in other words, noise, pauses, unclarity, etc., reveal two phenomena.

Firstly, mean proportional SCR responses go up during pauses but fall afterwards. Their number went up during pauses when interpreters did not work, possibly indicating increased stress waiting for the end of the pause. Such a reaction during waiting might be attributed to a emotional reaponse.

Secondly, mean proportional SCR responses go up during and after noise. When it comes to unclear sound (including poorly audible speakers' voices

and echo), we see that not only the number but also the amplitude of the peaks is significant, suggesting strong responses to deteriorated sound. The physiological reaction of the interpreters could be an indication of both an attentional and an emotional response.

Skin Conductance Responses correlate with different salient events interpreters encounter. Given that electrodermal measures respond both to cognitive and emotional engagement, they might be virtually impossible to disentangle without triangulating measures (Richardson et al., 2020). Thus, these salient effects found in this study deserve further exploration in controlled experiments.

References

Benedek, Mathias; Kaernbach, Christian (2010): A continuous measure of phasic electrodermal activity. In Journal of Neuroscience Methods 190 (1), pp. 80–91. DOI: 10.1016/j.jneumeth.2010.04.028.

Black, V. (1955). Laboratory versus field research in psychology and the social sciences. The British Journal for the Philosophy of Science 5(20), 319-330.

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust Multimodal Cognitive Load Measurement. Human-Computer Interaction Series.* Cham: Springer.

Critchley, H. D. (2002). Electrodermal responses: What happens in the brain. The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 8(2), 132–142. https://doi.org/10.1177/107385840200800209

Dawson, M. E., Schell, A. M., Filion, D. L., & Berntson, G. G. (2007). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), Handbook of Psychophysiology (pp. 157–181). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396.007

Richardson, D. C., Griffin, N. K., Zaki, L., Stephenson, A., Yan, J., Curry, T., Noble, R., Hogan, J., Skipper, J. I., & Devlin, J. T. (2020). Engagement in video and audio narratives: contrasting self-report and physiological measures. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-68253-2

Rojo Lopez, A. M. and Korpal, P (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. Linguistica Antverpiensia 19, 191-217.

Russo, M., Bendazzoli, C., and Defrancq, B. (Eds.)(2018). Making way in corpus-based interpreting studies. Singapore: Springer.

Seeber, K.G. (2020). Distance interpreting: Mapping the landscape. In Ahrens, B., Beaton-Thome, M., Krein-Kühle, M., Krüger, R., Link, L., & Wienen, U. (Eds.): Interdependenzen und Innovationen (pp. 123-172). Berlin: Frank & Timme.

Seeber, K., & Fox, B. (2022). Distance conference interpreting. In Routledge handbook of conference interpreting (pp. 491-507). Routledge. https://doi.org/10.4324/9780429297878-43

Sequeira, H., Hot, P., Silvert, L., & Delplanque, S. (2009). Electrical autonomic correlates of emotion. International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology, 71(1), 50–56. https://doi.org/10.1016/j.ijpsycho.2008.07.009

UNESCO. (1976). A teleconference experiment: A report on the experimental use of the Symphoniesatellite to link Unesco Headquarters in Paris with the Conference Centre in Nairobi.

••••

Tatiana Serbina
RWTH Aachen University
tatiana.serbina@ifaar.rwth-aachen.de

Mario Bisiada
Universitat Pompeu Fabra
mario.bisiada@upf.edu

Stella Neumann
RWTH Aachen University
stella.neumann@ifaar.rwth-aachen.de

Linguistic variation across different groups of translated and non-translated texts: Combined effect and individual contributions of lexico-grammatical features

The present study aims at investigating linguistic similarities and differences between different groups of texts, namely English originals, German translations and German originals. Additional sources of linguistic variation are introduced by different registers and revision stages included in the data. While the Geometric Multivariate Analysis (GMA) (e.g. Diwersy et al., 2014; Evert and Neumann, 2017), used in the study, emphasizes the combined effect of selected linguistic features, special attention will be paid to the analysis of individual contributions of features.

Previous studies have shown that the variables of language, register, translation status as well as editorial intervention have an effect on the lexico-grammatical features present in the texts (e.g. Delaere, 2015; Evert and Neumann, 2017; Bisiada, 2019; Serbina et al., 2021). Taking this into account, the present multivariate analysis combines these four variables corresponding to the different text groups present in the data sample. The study is based on a broad range of lexico-grammatical features, first introduced in Neumann (2013), including frequency counts of individual parts of speech as well as of more complex syntactic patterns, such as passive voice or imperative mood.

The data set consists of texts from two corpora, namely the Harvard Business Corpus (Bisiada, 2018) and the CroCo Corpus (Hansen-Schirra et al.,

2012). The Harvard Business Corpus contains articles published in English in the Harvard Business Review, edited German translations of these articles published in the Harvard Business Manager as well as German translation manuscripts of the same articles. The latter represent translation versions submitted by the translators and forwarded to the publisher by the translation company. The articles could be classified as belonging to the register BUSINESS. The second part of the data set contains English and German originals as well as published German translations taken from the CroCo corpus. The texts belong to the registers of letters to shareholders (SHARE), popular-scientific literature (POPSCI), prepared speeches (SPEECH) and political essays (ESSAY). The multivariate analysis of texts involves visual inspection and linguistic interpretation of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

Preliminary results suggest that particularly the variables of register, language and translation status appear to play an important role in text distribution compared to the variable of editorial intervention. The first PCA dimension groups texts based on both the language and the translation status: while the English and German originals are grouped on the negative and positive sides of the axis, respectively, the German translations included in the data set are clustered closer to the center. This distribution is in line with the results reported in the study applying GMA to several registers of the CroCo corpus: Evert and Neumann (2017) attribute this distribution to the shining-through effect of translations. The variation along the second dimension could be linked to the explanatory variable of register. As noted by Serbina et al. (2021), BUSINESS and SHARE registers are separated most distinctly.

To facilitate linguistic interpretation of the results, the focus of this paper is on the discussion of feature weights, which reflect the contributions of the individual lexico-grammatical features to the overall distribution of texts. This analysis will be performed with the help of the interactive "Weights Viewer" (see online supplement of Neumann and Evert, 2021), which offers a visualization of feature weights in the form of box-plots for different text groups.

References

Bisiada, M. (2018). Editing nominalisations in English–German translation. *The Translator*, 24 (1), 35-49.

Bisiada, M. (2019). Translated language or edited language? A study of passive constructions in translation manuscripts and their published versions. *Across Languages and Cultures*, 20 (1), 35-56.

Delaere, I. (2015). *Do Translations Walk the Line? Visually Exploring Translated and Non-Translated Texts in Search of Norm Conformity.* Ph.D. thesis, Ghent.

Diwersy, S., Evert, S. & Neumann, S. (2014). A weakly supervised multivariate approach to the study of language variation. In B. Szmrecsanyi & B. Wälchli (Eds.), *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech.* Berlin: de Gruyter, 174-204.

Evert, S. & Neumann, S. (2017). The impact of translation direction on characteristics of translated texts: A multivariate analysis for English and German. In G. de Sutter, M.-A. Lefer & I. Delaere (Eds.), *Empirical Translation Studies: New Theoretical and Methodological Traditions.* Berlin: de Gruyter, 47-80.

Hansen-Schirra, S., Neumann, S. & Steiner, E. (2012). *Cross-linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German.* Berlin: de Gruyter.

Neumann, S. (2013). *Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German.* Berlin: de Gruyter.

Neumann, S. & Evert, S. (2021). A register variation perspective on varieties of English. In E. Seoane & D. Biber (Eds.), *Corpus-based Approaches to Register Variation.* Amsterdam: John Benjamins, 143-178. Online supplement: https://www.stephanie-evert.de/PUB/NeumannEvert2021/.

Serbina, T., Bisiada, M. & Neumann, S. (2021). Linguistic profiles of translation manuscripts and edited translations. In *Proceedings for the First Workshop on Modelling Translation: Translatology in the Digital Age*, 34-45. Association for Computational Linguistics.

••••

Farnoosh Shamsian
Universität Leipzig, Germany
farnoosh.shamsian@uni-leipzig.de

Gregory Crane
Tufts University, United States
gregory.crane@tufts.edu

Corpus-based translation training: Enhancing translations from a historical language

Our work explores different aspects in which digital methods have changed our ability to access source texts in historical or low-resource languages. We look into how using parallel corpora and digital annotations have made it possible for Persian speakers to not only engage directly with the Ancient Greek texts, but to improve and refine indirect translations already available in Persian.

Although Persian comes from a very different cultural sphere, it has deep ties to ancient Greek as a result of years of cultural exchange. Many of the resources about the history of Persian-speaking countries and their cultural heritage are originally in Greek (Brosius, 2013). Meanwhile, the possibility of direct engagement with the Greek text is almost non-existent in the educational system; and direct Persian translations of Greek texts are extremely rare. However, indirect translations mainly through English, French and German are very common; and students and scholars from different backgrounds have been engaging with Greek literature through indirect translations instead of the original sources.

By understanding and incorporating different annotations, Persian speakers have the possibility to expand their knowledge of the source text without relying solely upon a mediating translation. They could cross linguistic boundaries by immediately engaging with the Greek texts and developing a more dynamic connection between the original source and its translation (Crane, 2019).

We report on the results of a 30-hour beginner-level Ancient Greek course in Persian based on book one of the Iliad aiming to improve learners' translation competence by integrating the emerging methods and tools from computational and data science (Mugelli, 2021; Mambrini, 2016). The course was

developed through a frequency-based syllabus resulting from the analysis of a heavily annotated corpus. After learning the basics of grammar, learners were introduced to different types of annotation including translation alignment, part of speech tags, treebanks, and other grammatical explanations. Translation alignment was often used as a pedagogical tool throughout the course (Palladino, 2020). Using linguistic scaffolding, learners were able to edit and refine the indirect translations already available in Persian and produce translations that were superior and more precise.

Moreover, we present a new methodology based on translation alignment for measuring improvements made in the indirect translations by the learners. To observe how close the translations were to each other and to the original text, we extracted the translation pairs and compared the ratios of 1-1, 1-N, N-1, and N-N pairs among different translations. When aligned using a guideline, we see noticeable differences in trends between the original indirect translation and the final refined version edited by the learners. In addition to the statistical comparisons, we look into the cultural aspect of translating the Iliad to Persian. Translation and translation alignment are regarded as a medium between cultures through which the learners can reflect on the complexities of a Greek epic while exploring the similarities to Shahnameh, the epic of Persian kings; a connection that at times has become evident in learners' choice of words.

We will conclude with the results of a translation course starting in Spring 2022 focused on producing a new translation of Plato's Crito. In this course, beginner and intermediate learners edit the available indirect translations of Crito by applying the same methods used in the Homeric Greek course to develop a parallel corpus with word-level translation alignment. Following the same methodology, we explore and assess their ability to enhance available translations, provide a word-level translation alignment and add further grammatical annotations to the Greek text.

References

Brosius, M. (2013). *Greek Sources on Achaemenid Iran.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199733309.013.0024

Crane, G. (2019). Beyond Translation: Language Hacking and Philology. *Harvard Data Science Review, 1*(2). https://doi.org/10.1162/99608f92.282ad764

Mambrini, F. (2016). The Ancient Greek Dependency Treebank: Linguistic Annotation in a Teaching Environment. *Ubiquity Press*. https://doi.org/10.5334/bat.f

Mugelli, G., Re, G., Boschetti, F., & Taddei, A. (2021). Learning Greek and Latin Through Digital Annotation: The EuporiaEDU System. *Teaching Classics in the Digital Age*, 25–36. https://doi.org/10.38072/2703-0784/p18

Palladino, C. (2020). Reading Texts in Digital Environments: Applications of Translation Alignment for Classical Language Learning. *The Journal of Interactive Technology and Pedagogy*. https://jitp.commons.gc.cuny.edu/reading-texts-in-digital-environments-applications-of-translation-alignment-for-classical-language-learning/

••••

Sabine Schmitz
Université Libre de Bruxelles, Belgium
sabine.schmitz@ulb.be

Exploring actual and perceived L2 translation difficulties experienced by student translators (French-German)

Since the end of the 1990s, translation studies have shown an increasing interest in translation into a second language, though it is still an under-researched topic (Hunziker Heeb 2018). More recently, several authors (Pavlović 2010; Ferreira et al. 2016) have investigated the differences between L1 and L2 translation and the difficulties underlying both processes. However, few (like Duběda, Mraček & Obdržálková 2018) have attempted to correlate translation errors with student translators' perceptions on the linguistic and translational challenges that have to be overcome.

Connecting the fields of corpus-based empirical translation studies, foreign language learning and translation teaching, this research tries to find out which aspects actually trigger mistakes in French-to-German student translations and which aspects are experienced as challenging by the students themselves. The project in addition aims to identify which strategies students use to overcome these perceived difficulties. To this end, three research instruments were designed. The first instrument is an error analysis grid used to assess a corpus of 95 translations produced by bachelor and master students, including different source text types (newspaper articles, university information brochures, commercial letters). Error categories include grammar and translation errors such as content distortion or pragmatic inadequacies. The second instrument is a questionnaire which aims to establish student profiles (with questions about their linguistic and academic background) and to elicit information about what they experienced as challenging (e.g. text length or register). The third instrument is an interview to give students the possibility to express themselves more freely. They were asked for instance to indicate which parts of the source text they perceived as the most difficult and to identify whether this had to do with grammar, style, lexis or pragmatic aspects. They were also asked to comment on the strategies and techniques they used to overcome perceived difficulties.

This research is driven by the hypothesis that translation students are not necessarily aware of the typical pitfalls and the aspects that need more focussed attention. Both bachelor and master students indeed consistently report that they struggle in equal measure with grammar and syntax on the one hand, and with lexis on the other hand. However, grammar and syntax errors account for 58.5% of all errors, whereas lexical errors represent "only" 13.2% of all errors. In other words, lexis is perceived as difficult as grammar, but causes far fewer errors. However, these errors tend to have a greater impact on the translation product, as they often have a direct influence on the readability of the target text.

Other examples of discrepancies between actual and perceived difficulties can be found at the syntactic level. Among the students who report they struggle with grammar and syntax in general, 74.3% find that syntax is an issue when translating into German, whereas 60.0% point to declension as one of the most difficult aspects. However, the corpus contains far more declension errors (35.8% of all grammatical errors) than syntactic errors (13.3%). Even more significantly, only two students reported they were not able to fully understand the source text. Content distortion – i.e. instances "when the source-text content is misrepresented in the translation" (Granger&Lefer 2021:11) – accounts for 12.1% of all errors, thus nearly equalling the number of syntactic errors. One explanation could be that many occurrences of distortion errors are due to wrong lexical choices in the L2 rather than a defective understanding of the source text, meaning that student translators are not aware of the actual problem.

From the interviews it emerges that only 1 in 2 students were able to find an appropriate strategy to solve a perceived problem. This too appears to support the hypothesis that a significant number of students may not be aware of the actual difficulties underlying L2 translation.

This corpus-driven research aims to provide new insights into L2 translation pedagogy. If translation trainers are given the tools to help students to correctly identify a problem and find the appropriate strategies to overcome it, the quality of their translations will improve significantly. Examples extracted from the corpus will allow translation trainers to design targeted exercises for specific student profiles which emerge from our error analyses.

Further research on parallel corpora of multiple language pairs would allow to establish whether our observations and conclusions could be extrapolated beyond French-to-German L2 translation.

References

Duběda, T., D. Mraček & V. Obdržálková (2018). *Překlad do nemateřského jazyka: Fakta, otázky, perspektivy.* Praha: Karolinum.

Ferreira, A., J.W. Schwieter, A. Gottardo & J. Jones (2016). Cognitive Effort in Direct and Inverse Translation Performance: Insight from Eye-Tracking Technology. *Cad. Trad., Florianópolis,* 36 (3), 60-80. https://doi.org/10.5007/2175-7968.2016v36n3p60

Granger, S. & M.-A. Lefer (2021). *Translation-oriented Annotation System manual. Version 2.0.* Louvain-la-Neuve : CECL Papers 3.

Hunziker Heeb, A. (2018). Professional translators' self-concepts and directionality: indications from translation process research. https://digital-collection.zhaw.ch/bitstream/11475/7349/2/Hunziker_Heeb_article.pdf.

Pavlović, N. (2010). What were they thinking?! Students' decision making in L1 and L2 translation processes. *Hermes, Journal of Language and Communication Studies,* 44, 63-87. http://download2.hermes.asb.dk/archive/download/Hermes-44-pavlovic.pdf.

••••

Ashon Song
Beijing Foreign Studies University, China
20672004@bfsu.edu.cn

Deforeignization of China's Political Translation: A Diachronic Perspective on Untypical Collocations

Untypical collocations, as indicated by the hypothesis of untypical lexical patterning in translations (Jantunen 2004), can be frequently found in translated languages. In China's political translation, for example, due to the lack of equivalence in the target language, untypical collocation is often the only option if one wishes to retain the elements of the source language (e.g. "spiritual civilization" for the Chinese concept "精神文明").

What is an untypical collocation? Despite the various definitions provided in previous studies, a clear-cut concept of "untypical collocation" is out of the question, because the judgement about being "untypical" is ultimately subjective. Therefore, inspired by Kenny's (2001) method, with reference to previous studies on the definition of collocation, we employed an operational definition of untypical collocation in the present study: a word string is defined as an untypical collocation when it is 1) grammatically constructed (Kjellmer 1984); 2) categorized as restricted combination (Aisenstadt 1979); 3) not found in BNC(1994) corpus; 4) judged as "unusual" by more than 3 native English speakers; and 5) co-occur with mutual expectancy greater than chance (Sinclair 1991), with the mutual information (MI) larger than 3 (Hunston 2002).

Based on this definition, using corpus tools (Antconc 3.5.9), we extracted 18 of the most representative untypical collocations from a corpus of the English translation and the original Chinese of the Government Work Report of the State Council of PRC (2001- 2020). The reason we choose the government work report as the research object is because the its translation is done through collective efforts, and it is one of the most read Chinese political text both in China and throughout the world. The size of the parallel corpus is 294,799 words in English, 374,210 words in Chinese, and 13,439 sentence pairs in total. We identified the source Chinese expressions of the 18 untypical English collocations, and observed the frequencies of both the Chinese and English expressions throughout the 20 years. Statistical results suggest

a "deforeignization trend" of the untypical collocations in recent years: most of the untypical collocations are no longer used in recent years, and less and less Chinese expressions are now translated into "foreignized" collocations. As is also reflected in the arguments of many official translators, more stress is now laid on the fluency and naturalness of the translations, compared with prevalent arguments for "correctness" or "accuracy" in the early years.

However, while untypical collocation may obscure the language and make one flinch from reading it, it retains the elements of the source language, and thus can be refreshing and impressive to foreign readers. Therefore, to use untypical collocations or not, in Toury's (1995) term, is a trade-off between adequacy and acceptance. After a closer examination, we divided the 18 untypical collocations into 4 categories: China-specific Noun Collocations, General Noun Collocations, Verb-Object Collocations, and Modifier Collocations. According to the statistics, only China-specific Noun Collocations are still used in recent years (e.g., "Moderately Prosperous Society" for "小康社会"), and the rest of the categories are all "deforeignized" and replaced by those more common and natural expressions (e.g., "leading cadres" are replaced by "government officials" for "领导干部"). It appears that todays Chinese government are adopting a strategy to "talk to him in his language" in its political text translation, which epitomized China's ever-increasing participation in the international community, and responded to President Xi Jinping's call to "tell Chinese stories in a reader-friendly manner".

References

Aisenstadt, E. (1979). Collocability restrictions in dictionaries. ITL-International Journal of Applied Linguistics, 45(1), 71-74.

Jantunen, J. H. (2004). Untypical patterns in translations. Translation universals: Do they exist, 48, 101-126.

Kenny, D. (2014). Lexis and creativity in translation: A corpus-based approach. Routledge.

Kjellmer, G. (1984). Some thoughts on collocational distinctiveness in recent developments in the use of computer corpora in English language research. Costerus, 45, 163-171.

Sinclair, J., & Sinclair, L. (1991). Corpus, concordance, collocation. Oxford University Press, USA.

Toury, G. (1995). Descriptive translation studies and beyond (Vol. 4). Amsterdam: J. Benjamins

Hunston S. (2002). Corpora in Applied Linguistics. Stuttgart: Ernst Klett Sprachen.

....

Vilelmini Sosoni
Ionian University
sosoni@ionio.gr

Maria Stasimioti
Ionian University
stasimioti@ionio.gr

Investigating quality and effort in direct and inverse post-editing

Although inverse translation, i.e., translation from a mother language (L1) to a non-native language (L2), has been broadly practiced around the world (Ferreira & Schwieter, 2017), it has received much less attention in previous studies than direct translation, i.e., translation from a non-native language (L2) to a mother language (L1) (ditto). Moreover, it has caused controversy among translation practitioners and has divided scholars as to its acceptability and feasibility. For instance, Ladmiral (1979), Newmark (1981), García-Yebra (1983) and Kelly et al. (2003) disfavour inverse translation, while Pokorn (2005) and Mraček (2018), among others, point to the fact that the translation process is not necessarily more effective into L1 and that translation quality is significantly dependent upon other factors, including "the individual capacities of the particular translator, his/her translational strategy, and his/her knowledge of the source and target cultures" (Pokorn, 2005: 121). Yet, aside from its level of quality, inverse translation has been found to be more time-consuming, effortful and cognitively demanding than direct translation (Buchweitz & Alves 2006; Pavlovic & Jensen, 2009; Fonseca & Barbosa, 2015; Hunziker Heeb, 2015; Ferreira 2014; Ferreira et al., 2016; Feng, 2017).

Directionality has received even less attention in the case of post-editing (PE) which has been increasingly gaining ground (O'Brien & Simard, 2014; Vieira et al., 2019), especially following the advent of Neural Machine Translation (NMT) and the improvement of the quality of the Machine Translation (MT) output especially at the level of fluency (Castilho et al., 2017a and 2017b). The few studies that have been carried out to date compare the quality of the post-edited text in L1 with the quality of the post-edited text in L2 (Garcia 2011; Sánchez-Gijón and Torres-Hostench, 2014; Toledo Báez,

2018) or the effort expended when carrying out PE in L1 and L2 (Igor et al., 2017). The studies' findings conclude that directionality does not imply differences in quality, although they do indicate that PE in L2 may demand more temporal effort than PE in L1 and also more cognitive effort especially in the area where the L2 text is produced.

Following the increasing use of multiple methods within the same study (Shreve & Angelone, 2010; Alves et al, 2010; Xiao, 2013; Halverson, 2017), this paper adopts a mixed-methods approach and integrates both product- and process-oriented perspectives to investigate inverse PE and its effect on translators and on the translation product. In particular, it seeks to compare the cognitive, temporal and technical effort required for full PE of NMT output in L1 with the effort required for full PE of NMT output in L2, focusing on the English-Greek language pair for which there are no related studies to date. In particular, eye-tracking and keystroke logging data obtained using Translog-II connected to a Tobii X2-60 eye tracker are used to measure the effort expended by translators while carrying out full PE of NMT output. Questionnaires are also used to capture the translators' attitudes and perceptions and a fine-grained human error analysis evaluates the quality of the final translation product.

The study's findings indicate that inverse PE is less demanding in terms of cognitive, technical, and temporal effort than direct PE. They also reveal that in both direct and inverse PE much of the activity involved in the PE task takes place in the TT area given. However, in the case of direct PE a higher visual attention on the ST area is observed presumably due to the more careful reading and processing of the ST that is written in a foreign language, i.e., English, as well as due to the translators' need to not only feed their brain with input for meaning construction but also to ensure while typing that the TT conveys the meaning of the ST.

Another interesting finding that emerges from the study is that directionality does not imply differences in quality. In particular, during the evaluation of the correctness and necessity of the edits performed during direct and inverse PE, the percentage of correct edits (i.e., the sum of correct and necessary edits as well as correct and necessary no edits) was almost the same in the case of inverse and direct PE.

References

Alves, F., Pagano, A. Neumann, S., Steiner, E, & Hansen-Schirra, S. (2010). Translation units and grammatical shifts. Towards an integration of product- and process-based translation research. In Shreve, G. M., & Angelone, E. (Eds). *Translation and Cognition* (pp. 109-142). Amsterdam: John Benjamins.

Buchweitz, A., & Alves. F. (2006). Cognitive adaptation in translation: an interface between language direction, time, and recursiveness in target text production. *Letras de Hoje, 41*(2), 241–272.

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017a). Is neural machine translation the new state of the art? *Prague Bulletin of Mathematical Linguistics,* 108, 109–120.

Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, Y., Lohar, P., Way, A., Miceli Barone, A., & Gialama, M. (2017b). A Comparative quality evaluation of PBSMT and NMT using professional translators. In *Proceedings of Machine Translation Summit XVI*, vol.1: Research Track. Nagoya, Japan, 18-22 September 2017. 16th Machine Translation Summit, pp. 116–131.

Feng, J. (2017). Comparing Cognitive Load in L1 and L2 Translation: Evidence from Eye-Tracking. *Foreign Languages in China* (《中国外语》), (78), 79-91.

Ferreira, A., & Schwieter, J. W. (2017). Directionality in Translation. In Schwieter, J. W. & Ferreira, A. (Eds.), *The Handbook of Translation and Cognition* (pp. 90-105). Hoboken, NJ: Wiley-Blackwell.

Ferreira, A., Schwieter, J. W., Gottardo, A, & Jones, J. (2016). Cognitive effort in direct and inverse translation performance: insight from eye-tracking technology. *Cadernos de Tradução*, *36*(3), 60-80.

Ferreira, A. (2014). Analyzing Recursiveness Patterns and Retrospective Protocols of Professional Translators in L1 and L2 Translation Tasks. *Translation and Interpreting Studies*, *9*(1), 109-127.

Fonseca, D. L., & Barbosa, N. (2015). Directionality in translation: Investigating prototypical patterns in editing procedures. *Translation & Interpreting,* 7, 111-125.

Garcia, I. (2011). Translating by post-editing: is it the way forward?. *Machine Translation*, *25*(3), 217- 237.

García-Yebra, V. (1983) *En Torno a la Traducción*. Madrid: Gredos.

Halverson, S. L. (2017). Multimethod approaches. In Schwieter, J. W. & Ferreira, A. (Eds.), *The Handbook of Translation and Cognition* (pp. 195-212). Hoboken, NJ: Wiley-Blackwell.

Hunziker Heeb, A. (2015). Does Professional Translation into L2 Involve More Effort than into L1? Oral communication at 4th Translation Process Research Workshop. Las Palmas, January 15-17.

Igor, L.S., Alves, F., Schmaltz, M., Pagano, A., Wong, D., Chao, L., Leal, A.L., Quaresma, P., Garcia, C., & Da Silva, G.E. (2017). Translation, Post-Editing and Directionality: A Study of Effort in the Chinese-Portuguese Language Pair. In Jakobsen, A.L., & Mesa-Lao, B. (Eds.). *Translation in Transition. Between cognition, computing and technology* (pp. 91-117). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Kelly, D., Martin, A., Nobs, M. L., Sanchez, D., & Way, C. (2003) Reflexiones en Torno a Algunos Conceptos Básicos. In Kelly, D., Martin, A., Nobs, M. L., Sanchez, D., & Way, C. (Eds) *La direccionalidad en Traducción e Interpretación. Perspectivas Teóricas, Profesionales y Didácticas* (pp. 33-41). Granada: Atrio.

Krings, H. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent: Kent State University Press.

Ladmiral, J. R.. (1979). *Théorèmes pour la Traduction*. Paris: Didier

Mráček, D. (2018) Inverse translation: the more challenging direction. *Linguistica Pragensia*, 28, 202- 221.

Newmark, P. (1981). *Approaches to Translation*. Oxford: Pergamon

O'Brien, S. (2002). Teaching post-editing: A proposal for course content. In: *EAMT 2002 Proceedings of the 6th annual conference of the European association for machine translation*. Manchester, UK, 14- 15 May 2002. European Association for Machine Translation, pp. 99–106.

O'Brien, S., & Simard, M. (2014). Introduction to special issue on postediting. *Machine Translation 28*(3), 159–164

Pavlović, N., & Jensen. K. T. H. (2009) Eye tracking translation directionality. In Pym, A., & Perekrestenko, A. (Eds) *Translation Research Projects* 2 (pp. 101-119). Tarragona: Universitat Rovira i Virgili.

Pokorn, N. (2005). *Challenging the Traditional Axioms: Translation into a Non-Mother Tongue.* Amsterdam: John Benjamins

Sánchez-Gijón, P. & Torres-Hostench, O. (2014). MT Post-editing into the Mother Tongue or into a Foreign Language? Spanish-to-English MT translation output post- edited by translation trainees. In O'Brien, S., Simard,

M. & Specia, L. (Eds) *Proceedings of the Third Workshop on Post-editing Techniques and Practices (WPTP-3)*: The 11th Conference of the Association for Machine Translation in the Americas: October 22-26, 2014: Vancouver, BC Canada. AMTA, pp. 15-19.

Shreve, G. & Angelone, E. (Eds). (2010). *Translation and Cognition*. Amsterdam and Philadelphia: John Benjamins.

Toledo Báez, M. C. (2018). Machine Translation and Post-editing: Impact of Training and Directionality on Quality and Productivity. *Revista Tradumàtica. Technologies de la Traducció*, 16, 24- 34.

Vieira, L.N., Alonso, E., Bywood, L. (2019) Introduction: Post-editing in practice – Process, product and networks, *Jostrans: The Journal of Specialised Translation*, 31, 2-13.

Xiao, K. (2013) Cognitive approaches to translation process: Current trends, challenges and future development. In Awang, R. A., Ghani, A. A. G. & Ayob L. A. (Eds) *Translator and Interpreter Education and Training: Innovation, Assessment and Recognition*. Kuala Lumpur: Malaysian Translators Association.

••••

Carola Strobl
University of Antwerp
carola.strobl@uantwerpen.be

Gert De Sutter
Ghent University
Gert.DeSutter@UGent.be

Jim J.J. Ureel
University of Antwerp
jim.ureel@uantwerpen.be

Sonia Vandepitte
Ghent University
Sonia.Vandepitte@UGent.be

A pilot analysis of co-reference in German–Dutch translation: Probing into (a)symmetry in the translation of cohesion.

Cohesion is crucial in writing and translation since it aids reader orientation. At the same time, cohesion is highly language specific: Each language has its own grammatical and lexical devices to connect sentences, paragraphs and ideas in texts and its own preferred ways to allow users to deploy cohesive devices in specific genres. Even languages such as German, Dutch and English, which share many lexical and grammatical devices that can be used to create cohesion, differ with regard to the preferred use of those devices (Kunz et al., 2021; Van de Velde, 2011). This language-specific variation makes cohesion an interesting case for the investigation of (a)symmetry in translation (Becher, 2011; Blum-Kulka, 1986): grammatical and lexical cohesion are intrinsically intertwined, while the preference for one of both is highly language- (and genre-) dependent.

Lexical repetition can be considered an overt, *explicit*, strategy to create cohesion, while grammatical cohesion, such as co-reference (e.g., personal or demonstrative pronouns), must be decoded by readers/listeners and can, thus, be considered a covert, *implicit*, strategy. In the translation process, co-reference means can be retained, omitted, added or replaced. Omission and addition always result in cohesion shifts (Blum-Kulka, 1986), whereas re-

placement, according to Blum-Kulka´s narrow definition, results in cohesion shifts only if co-reference devices in source texts are replaced with devices from other cohesive categories (e.g., lexical cohesion) in target texts.

While there is a considerable amount of research on differences between German and English with regard to cohesion (Kunz et al., 2021), and more specifically with regard to coreference strategies in translation (Grishina and Stede, 2015; Lapshinova-Koltunski et al., 2020), cohesion has received very little attention to date in contrastive research on the language pair German and Dutch. The only published study to date that examined explicitation and implicitation in a bidirectional German–Dutch translation corpus including cohesion is Van de Velde (2011), who found an asymmetrical predominance of explicitation for German to Dutch translation.

Our presentation aims to shed light on this under-researched domain. We will present the results of a pilot study into co-reference use in non-literary texts translated from German into Dutch and vice-versa. The research questions driving this study are: (1) Is the relationship between explicitation and implicitation of co-reference in German–Dutch and Dutch–German translation symmetrical or asymmetrical (Klaudy & Károly, 2005)? and, if it is asymmetrical, (2) Can cohesion shifts (Blum-Kulka, 1986) in German–Dutch and Dutch–German translation be attributed to so-called translation universals, or rather can they be attributed to differences in language-specific preferences for co-reference strategies?

The data used for this analysis stem from a new sub-corpus of the bidirectional parallel corpus for German–Dutch translation, PAND (Parallelcorpus Nederlands–Duits), collected at a Flemish university. This sub-corpus contains non-fictional texts of diverse origins within the cultural domain (e.g., museum texts, tourist leaflets), totalling 686,009 words. We manually analysed a selection of texts, totalling 24,072 words, along 13 co-reference devices pertaining to three categories: 4 devices of personal reference, 6 devices of demonstrative reference and 3 devices of comparative reference. Preliminary results reveal a non-significant higher relative frequency of co-reference devices in the Dutch original and translated texts than in their German counterparts. In addition, we found a significantly higher relative frequency of co-reference devices in original Dutch as compared with translated Dutch, and a significantly lower relative frequency of co-reference devices in original German as compared with translated German. These results suggest the tentative hypothesis of an asymmetrical relationship, that is,

supporting the asymmetry hypothesis. In addition, in-depth analyses of selected instances of changes in co-reference solutions reveal that they relate mostly to replacement *within* the cohesion paradigm of co-reference, i.e., exchanging one co-reference device with another. This means that few instances of actual cohesion shifts in the narrow definition have been attested, that is, co-reference translated by lexical repetition or vice-versa.

Based on the results of this small-scale pilot study, we cannot assume a significant tendency regarding explicitation or implicitation directionality. A clearer picture of the relationship between explicitation and implicitation in the translation of co-reference that might support the asymmetry hypothesis (Klaudy & Károly, 2005) would require the analysis of a larger number of texts. Based on examples of our in-depth study, we will discuss potential triggers for observed cohesion shifts, including language-specific preferences and so-called universal translation phenomena, such as shining-through and normalisation (Teich, 2003) and gravitational pull (Halverson, 2017).

References

Becher, V. (2011). When and why do translators add connectives? A corpus-based study. *Target: International Journal of Translation Studies,* 23 (1), 26–47.

Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies* (pp. 17–35). Tübingen: Gunter Narr.

Grishina, Y., & Stede, M. (2015). Knowledge-lean projection of coreference chains across languages. *Proceedings of the Eighth Workshop on Building and Using Comparable Corpora* (pp. 14–22), Beijing, China (online). Association for Computational Linguistics.

Halverson, S. (2017). Gravitational pull in translation: Testing a revised model. In G. De Sutter, M.-A. Lefer, & I. Delaere (Eds.), *Empirical translation studies: New methodological and theoretical traditions* (pp. 9–46). Berlin, Germany; Boston, Massachusetts: De Gruyter Mouton.

Klaudy, K. & Károly, K. (2005). Implicitation in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures,* 6 (1), 13–28.

Kunz, K., Menzel, K., Steiner, E., Lapshinova-Koltunski, E., & Martínez Martínez, J. M. (2021). *GECCo – German–English contrasts in cohesion. Insights*

from corpus-based studies of languages, registers and modes. Berlin, Germany; Boston, Massachusetts: De Gruyter Mouton.

Lapshinova-Koltunski, E., Krielke, M., & Hardmeier, C. (2020). Coreference Strategies in English-German Translation. *Proceedings of the Third Workshop on Computational Models of Reference, Anaphora and Coreference* (pp. 139–153), Barcelona, Spain (online). Association for Computational Linguistics.

Teich, E. (2003). *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts.* Berlin, Germany; Boston, Massachusetts: De Gruyter Mouton.

Van de Velde, M. (2011). Explizierung und Implizierung im Übersetzungspaar Deutsch–Niederländisch: Eine quantitative Untersuchung. In P. A. Schmitt, S. Herold, & A. Weilandt (Eds.), *Translationsforschung* (pp. 865–884). Frankfurt/Main: Peter Lang.

••••

Elina Symseridou
Aristotle University of Thessaloniki, Greece
e.symseridou@gmail.com

The use of Corpora in the Teaching of Specialised Translation: a Theoretical and Practical Approach

One of the key issues faced by translators and translation students of specialised texts is finding the equivalents of terms in L2 of the field in question. A greater challenge, however, is the formation of the textual environment with the appropriate collocations (adjectives, nouns, verbs) for those terms in the language for special purposes (LSP). The web offers the most convenient and immediate solution by providing access to updated language data presenting the terms in original contexts that help overcome the shortcomings of hard copy lexicographic resources.

Taking into account the importance of documentation skills and electronic-tools literacy in the training of translators of specialised texts, known as information mining and technological competences (EMT Expert Group, 2009), in this presentation I will provide an overview of the findings of my doctoral research focusing on the subject of ad hoc specialised corpora, i.e. the compilation of individualised text collections using WebBootCat (Baroni et al. 2006), a sophisticated automatic corpus-building tool, for the purposes of specialised translation and translator training. More specifically, the aim of this presentation is to report on the final results of a research that investigated:

1. the degree of integration of corpus use and corpus technology in translator training programs offered by two Greek higher education institutions,
2. the use of the Web as a source to collect linguistic data and a means for constructing corpora automatically with the WebBootCat tool, available within the Sketch Engine corpus analysis platform.

In order for the above goals to be achieved, the following will be carried out:

a. present the results of a research study conducted on MA and BA students who used WebBootCat for constructing an ad hoc specialised corpus to help them translate a specialised text from French and English into Greek on the mental health issue of schizophrenia,

b. present the key-points of an interview, based on a semi-structured questionnaire, in which students participated to examine their views and expectations on the use of technological tools in translator training,

c. propose an educational model based on the combined use of an automatic corpus-building tool (WebBootCat) and a translation memory (such as Trados).

Overall, my research shows that corpora and corpus technology are not included among the translation tools used in educational environments, as opposed to translation memories, which are a prerequisite for the labour market, and that students seemed to have little or no idea of what an ad hoc corpus is or how they could create one automatically, using an online tool, to serve specialised-translation needs.

Furthermore, as far as the degree of familiarisation with translation memories is concerned it results that students felt confident about the technical know-how but not about application in the translation practice and production, and they were keen on the inclusion of more tools in translation training.

In brief, the contribution of my doctoral thesis lies in:

- Indicating a significant research gap concerning the use of WebBootCat, an automatic corpus-building tool, and the Sketch Engine innovative corpus analysis tools in translator training.
- Proposing an educational model based on the combined use of an automatic corpus-building tool (WebBootCat) and a translation memory (such as Trados), thus, integrating dynamically and practically into translator training two modern computing tools to support the act of translating.

To conclude, this research lays the foundation for the exploitation of a sophisticated tool for automatically constructing corpora from relevant web pages, which in combination with the use of translation memories it can contribute to the improvement of the translator's competences (according to the translator competence profile developed by the European Master's in Translation) on linguistic, translational and technological level.

Keywords: translator training; language for special (LSP); ad hoc special-ised corpora; WebBootCat; specialised translation;

References

Book

Kiraly, D. (2000). *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice.* Routledge, NY.

Article in edited volume

Bernardini, S. (2000). 'Systematising serendipity: proposals for concordan-cing large corpora with language learners' In: Burnard, L. and McEne-ry, T. (eds.). *Rethinking Language Pedagogy from a Corpus Perspective.* Frankfurt: Peter Lang: 225-234.

Bernardini, S. and Castagnoli, S. (2008). 'Corpora for translator education and translation practice' In: E. Yuste Rodrigo *(ed). Topics in Language Re-sources for Translation and Localisation.* Amsterdam/Philadelphia: John Benjamins: 39-55.

Fletcher, W. (2004). 'Facilitating the compilation and dissemination of ad hoc web corpora' In: Guy, A., Bernardini, S. and Stewart, D. (eds.). *Corpora and Language Learners. Amsterdam and Philadelphia: John Benjamins, 271-300.*

Kübler, N. (2011). 'Working with Corpora for Translation Teaching' In: Fran-kenberg-Garcia, A., Flowerdew, L. and Aston, G. (eds.). *New Trends in Corpora and Language Learning. London: Bloomsbury: 62-80.*

Varantola, K. (2003). 'Translators and Disposable Corpora' In: Zanettin, F., Bernardini, S. and Stewart, D. (eds). *Corpora in Translator Education.* UK: St. Jerome Publishing.

Journal article

Frankenberg-Garcia, A. (2015). *Training translators to use corpora hands-on: challenges and reactions by a group of thirteen students at a UK Univer-sity.* Edinburgh University Press, Vol. 10, Issue 3: 351-380. [Available at: http://www.euppublishing.com/doi/abs/10.3366/cor.2015.0081].

Frérot, C. (2016). *Corpora and Corpus Technology for Translation Purposes in Professional and Academic Environments. Major Achievements and New Perspectives.* Cadernos de Tradução, On-line version ISSN 2175-7968. [Available at: http://www.scielo.br/scielo.php?script=sci_arttext&-pid=S2175-79682016000500036].

Kilgarriff, A. and Grefenstette, G. (2003). *Web as Corpus*. [Available at: https://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WA-CIntro.pdf]

Conference paper

Baroni, M. and Bernardini, S. (2004). *BootCat: Bootstraping corpora and terms from the web* [Available at: http://clic.cimec.unitn.it/marco/publications/lrec2004/bootcat_lrec_2004.pdf]

Baroni, M., Kilgarriff, A., Pomikálek, J. and Rychl, P. (2006). *WebBootCat: Instant Domain-specific Corpora to Support Human Translators*, EAMT 2006 [Available at: http://www.mt-archive.info/EAMT-2006-Baroni.pdf].

Picton, A., Fontanet, M., Pulitano, D. and Maradan, M. (2015). *Corpora in Translation: addressing the Gap between the Scholars' and the Translators' Point of View. CULT Conference, 26-29 May,* Alicante.

The EMT competence framework

EMT Expert Group (2009). *Competences for professional translators, experts in multilingual and multimedia communication.* [Available at: http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf].

••••

Denisa Šebestová
FF UK
denisa.sebestova@ff.cuni.cz

Dominika Kováříková
FF UK
kovarikova@korpus.cz

Translation equivalents of the Czech academic vocabulary list Akalex

The current trend of compiling academic vocabulary lists (AVLs) is extending from English to other languages, such as Portuguese (Batista et al. 2010), Swedish (Carlund et al. 2012), or Czech (Kováříková & Kovářík 2021). The national academic vocabulary lists serve different purposes from English ones, as the target audiences and their needs are different. Most of the English AVL target audience are second language speakers or students of academic writing courses and therefore the lists are usually used in the teaching/learning environment as learning aids. Although some of the national AVL users, especially in smaller languages such as Czech, are students as well, there is also a substantial group of experienced professional researchers writing academic texts in nonnative languages, mostly in English. One of the purposes of a national AVL is therefore providing a range of translation equivalents of each of the items (in addition to listing frequency, meanings, synonyms, and collocations). The aim of this study is to point out some issues of translating academic vocabulary items and also to present associated methodological challenges. The Czech academic vocabulary list Akalex (www.korpus.cz/akalex) was compiled in 2021 based on data from the SYN2015 and SYN2020 corpora of contemporary written Czech. The criteria for choosing the individual items are the frequency ratio in academic and non-academic texts, relative frequency in academic texts, number of disciplines in which the item appears and dispersion in the academic subcorpus. The list contains more than one thousand words and multi-word units, of which approximately 500 are nouns, 300 adjectives, 150 verbs, 60 adverbs and 100 other POS or MWUs (such as multi-word prepositions and others). The main challenge of translating academic vocabulary lies in listing only the relevant translation equivalents typical for academic texts. Instead of

verifying whether the equivalent is sufficiently frequent and evenly dispersed in academic disciplines (for such a task, a corpus of academic and non-academic English texts would be needed), the existing English AVLs can be consulted, for example the Academic Word List (Coxhead 2000), the Academic Keyword List (Paquot 2010) or the New Academic Vocabulary List (Gardner & Davies 2014). For this study, we focused on identifying the English translation equivalents of 154 Czech academic verbs. Most of the verbs belong to one of 8 semantic categories: research and thought activities, description and characterization of phenomena, classification of phenomena, relationships between phenomena, change or development, activities related to writing and description, modality, and reference (both intra-text and inter-text). Verbs were selected as the subject of this analysis since their translation poses a variety of interesting problems. First of all, the abovementioned relevance of translation equivalents is reflected in the choice of adequately formal verbs; the choice of verbs influences the overall style of the text, mainly whether it is perceived as formal or informal (cf. determine or identify vs. find out). Some of the verbs can only be translated by a non-verbal or a structurally different phrase (docházet k – there is, take place). International verbs present their own challenges – some of their seemingly straightforward equivalents can be used only in specific contexts (formulovat – to formulate, interpretovat – to interpret, demonstrovat – to demonstrate), several pairs of false friends were also detected (disponovat – to dispose of, redukovat – to reduce). Another issue related particularly to verbs is the identification of typical preposition or conjunction which typically combine with a given verb and the cross-linguistic differences in this respect (vyplývat, že – to result in). This comparison seems valuable since prepositions are a frequent source of errors (the translation equivalents of prepositions differ between languages both in terms of their usage and the scope of their meaning) (Klégr & Malá 2009). Similarly, any occurrence of language-specific multi-word units should be documented and commented on. Last but not least, it is important to detect and compare any specific grammatical characteristics of individual verbs, especially the ratio of active and passive voice or typical tense (especially with regard to verbal aspect). Translations into English are only one component of the Akalex academic vocabulary list; they are complemented with information on synonyms, collocations and in the future possibly even translations to other languages. Akalex thus provides a comprehensive tool helping scholars share the results of their research in adequate writing style.

References

Baptista, J. et al. (2010). P-AWL: Academic word list for Portuguese. *Proceedings of Computational Processing of the Portuguese Language PROPOR 2010*, Porto Alegre, Brazil.

Carlund, C. et al. (2012). An academic word list for Swedish - a support for language learners in higher education. *Proceedings of the SLTC 2012 workshop on NLP for CALL. Linköping Electronic Conference Proceedings,* 80, 20-27.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.

Gardner, D., Davies, M. (2014) A new academic vocabulary list. *Applied Linguistics*, 35, 305-327.

Klégr, A., & Malá, M. (2009). English equivalents of the most frequent Czech prepositions: a contrastive corpus-based study. *Proceedings of the Corpus Linguistics Conference, CL 2009*, Liverpool, 20-23 July 2009. http://ucrel.lancs.ac.uk/publications/cl2009/

Kováříková, D., Kovářík, O. (2021). *Akalex - Application for research of academic Czech*. Institute of the Czech National Corpus, Faculty of Arts, Charles University. Prague. Available at http://www.korpus.cz/akalex.

Paquot, M. (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. New York: Continuum.

Corpora

Křen, M. et al. (2015). *SYN2015: representative corpus of written Czech*. Institute of the Czech National Corpus, Faculty of Arts, Charles University. Prague. Available at http://www.korpus.cz.

Křen, M. et al. (2020). *SYN2020: representative corpus of written Czech*. Institute of the Czech National Corpus, Faculty of Arts, Charles University. Prague. Available at http://www.korpus.cz.

••••

Isabel Tello
Universidad Politécnica de Cartagena
isabel.tello@cud.upct.es

The translation of diminutives into Spanish: testing the Unique Items Hypothesis with COVALT corpus

Using diminutive suffixes is very common in Spanish. However, translators tend to underuse them when translating from languages where diminutives are scarcer or less used. In this paper, we will try to find out if this is the case in literary texts translated from English and French into Spanish. To this aim, we will use the theoretical framework proposed by Halverson's Gravitational Pull Hypothesis (GPH) (2003, 2010, 2017) and Tirkkonen-Condit's Unique Items Hypothesis (UIH) (2004). The idea underlying the UIH is that target language typical items are under-represented in translation, not over-represented as suggested by Baker (1993). The GPH, Halverson's cognitive attempt to explain the characteristics of translated language, identified three causes of translational effects (2010): 1) Patterns of prototypicality of target language items (TL), which would lead to over-representation of the TL item in the translation; 2) Conceptual structures or the representation of the source language (SL) item, leading to over-representation as well; and 3) Patterns of connectivity, which reflect relationships between the SL and the TL and would lead to over-representation or under-representation.

The research questions for this study are:

RQ1: Do Spanish translations from English and French use fewer diminutive suffixes than Spanish non-translated texts?

RQ 2: If so, are frequency differences between translations and non-translations in Spanish due to source language influence?

Diminutives behave differently in the three languages of the study. English and French diminutives follow a word-formational pattern: English *-let* and *-y* as in *booklet* and *kitty* (Sicherl, 2018), and derivative affixes in French such as *-eau/elle, -et/ette, -ot/otte, -in/ine, -on* and *-illon* (Klett, 2015: 176). However, these suffixes are not commonly used in any of these languages. Although diminutives in French are a bit more productive than in English, both languages prefer anteposition of *little/small* and *petit* respectively. Conversely, in Spanish the use of diminutives is very frequent, and we find very

productive suffixes such as *-ito/ita or -illo/illa*, among others (Lázaro, 1999). Considering this, and even if English and French have diminutives as well, diminutives for English-Spanish (EN-ES) and French-Spanish (FR-ES) will be considered unique items in this study.

Data analysis will be carried out using COVALT, which is a multilingual parallel corpus of narrative works originally written in English, French and German and translated into Catalan and Spanish. It is also a comparable corpus since it includes two sub-corpora of Catalan and Spanish non-translated works. Corpus data analysis will use Corpus Query Processor (CQP).

The study will try to validate the following hypotheses:

1. Diminutive suffixes will be under-represented (Halverson's factor 3 and UIH) in translations from English and French into Spanish, as compared to Spanish non-translations.
2. Under-representation will be less pronounced in translations from French, where the range of diminutive suffixes is broader than in English and they are somewhat more productive.

To answer RQ 1, diminutive suffixes -ito/ita/itos/itas and -illo/illa/illos/illas in Spanish non-translated texts (ES-OR) and Spanish translations from English (EN-ES) and French (FR-ES) were extracted. Since query matches were too many for manual analysis, they were randomly thinned. After manually sifting of the three sets of occurrences to remove false positives, results were compared and quantified. Query results needed to be normalised as the three sub-corpora do not have the same size. Table 1 shows the results of the quantitative analysis:

RESULTS	CORPUS SIZE	QUERY MATCHES	PROPER DIMINUTIVES (out of 500/700)	NORMALISED FREQUENCY PER 1,000 WORDS
TRANSLATIONS FROM ENGLISH	1,122,299	4,046	184	1.33
TRANSLATIONS FROM FRENCH	565,481	2,149	221	1.68
NON-TRANSLATIONS	4,170,178	16,291	317	1.77

Table 1. Frequency of occurrence in COVALT of the diminutive suffixes *-ito/ita/itos/itas* and *-illo/illa/illos/illas.*

Results showed that hypothesis 1 can be validated as there is under-representation of translated diminutives from English and French, which is consistent with Halverson's factor 3 (patterns of connectivity) and with UIH (under-representation of target language typical items when they do not have a direct counterpart in the ST). Hypothesis 2 is validated as well: under-representation is less pronounced in FR-ES due to French somehow more productive diminutive suffixes. These results can answer RQ1.

To answer RQ2, we analysed English and French triggers which had led to the translations. RQ2 aimed at finding out whether frequency differences between translations and non-translations in Spanish were due to source language influence.

After categorisation of possible triggers, table 2 summarises trigger results:

TRIGGERS	ENGLISH		FRENCH	
A. NO DIMINUTIVENESS	91	48.7%	56	25.3%
B. ST EXPLICIT DIMINUTIVENESS (little/petit and -let/-y/-et/-on, etc.)	82	43.9%	156	70.6%
C. IMPLICIT/INHERENT DIMINUTIVENESS	11	5.9%	9	4.1%
	187		221	

Table 2. Results of EN and FR triggers for diminutive suffixes *-ito/ita/itos/itas* and *-illo/illa/illos/illas.*

Trigger analysis in the corpus showed fewer diminutive triggers (including *little*/*small* and *-let*/*-y*) in English than in French. This supports the results of quantitative analysis above and may answer RQ2: as there are more explicit triggers (*petit*/*-on*/*-et*) in French, translators may have been more prone to use diminutives in their texts; in English, however, there are more instances of no diminutive triggers, which may have led to a less frequent use of diminutives in Spanish translations.

References

Halverson, S. (2003). The cognitive basis of translation universals. *Target,* 15(2), 197-241.

Halverson, S. (2010). Cognitive translation studies: Development in theory and method. In G. M. Shreve & E. Angelone (Eds.) *Translation and Cognition.* Amsterdam & Philadelphia: John Benjamins, 349-369.

Halverson, S. (2017). Gravitational pull in translation. Testing a revised model. In G. De Sutter, M. A. Lefer & I. Delaere (Eds.) *Empirical Translation Studies.* Berlin: De Gruyter Mouton, 9-45.

Hareide, L. (2017a). The translation of formal source-language lacunas. An empirical study of the Over-representation of Target-Language Specific Features and the Unique Items Hypothesis. In M. Ji, M. Oakes, L. Defeng, & L. Hareide (Eds.) *Corpus Methodologies Explained. An Empirical Approach to Translation Studies.* London & New York: Routledge, 137-187.

Hareide, L. (2017b). Is there gravitational pull in translation? A corpus-based test of the Gravitational Pull Hypothesis on the language pairs Norwegian-Spanish and English-Spanish. In M. Ji, M. Oakes, L. Defeng, & Lidun Hareide (Eds.) *Corpus Methodologies Explained. An Empirical Approach to Translation Studies.* London and New York: Routledge, 188-231.

Klett, E. (2015). El diminutivo en español y francés: rasgos transculturales y translingüísticos para su enseñanza. El hilo de la fábula. *Revista anual del Centro de Estudios Comparados,* (15), 173-185.

Lázaro Carreter, F. (1999). *Diccionario de términos filológicos.* Tercera edición corregida. Madrid: Gredos.

Marco, J. (2019). Living with parallel corpora. The potentials and limitations of their use in translation research. In I. Doval Reixa & M. T. Sánchez Nieto (eds.) *Parallel Corpora for Contrastive and Translation Studies: New Resources and Applications.* Amsterdam & Philadelphia: John Benjamins, 39-56.

Marco, J. (2021). Testing the Gravitational Pull Hypothesis on modal verbs expressing obligation and necessity in Catalan through the COVALT corpus. In M. Bisiada (ed*.) Empirical Studies in Translation and Discourse.* Berlin: Language Science Press. 27–52.

Marco, J. & Oster, U. (12-14 September 2018). The gravitational pull of diminutives in Catalan translated and non-translated texts. Using Corpora in Contrastive and Translation Studies Conference (5th edition). Louvain-la-Neuve, Belgium.

Oster, U. (2020). "Sobrerrepresentación del adjetivo antepuesto en textos traducidos: ¿realidad o prejuicio?". En Recio Ariza, M.ª Ángeles et al. (Eds). *Del texto a la traducción. Estudios en homenaje a Pilar Elena*. Granada: Comares. 115-132.

Real Academia Española (2010). *Nueva Gramática de la Lengua Española. Manual.* 1ª ed. Espasa.

Tirkkonen-Condit, S. (2004). Unique items – over- or underrepresented in translated language? In Anna Mauranen & Pekka Kujamäki (Eds.), *Translation universals. Do they exist?*, 177-184. Amsterdam: John Benjamins.

••••

Lise Volkart
University of Geneva, Switzerland
lise.volkart@unige.ch

Pierrette Bouillon
University of Geneva, Switzerland
pierrette.bouillon@unige.ch

Building a corpus for the study of post-editese in professional context: methodology and challenges

1. Motivation

Post-editese is a relatively recent concept, echoing the translationese phenomena, that was first mentioned by Daems et al. in 2017. It can be defined "as the expected unique characteristics of a PE [post-edited] text that set it apart from a translated text" (Daems et al., 2017). In recent years, several scholars, (Daems et al. (2017), Toral (2019), Castilho et al. (2019) and Castilho & Resende (2022)) have tried to demonstrate the existence of post-editese and to identify its characteristics. Most of them investigated post-editese using automatic metrics such as type/token ratio (TTR), ratio of content words, length ratio, etc. on parallel corpora (i.e. a single source text translated once by a human and once by a machine translation [MT] system and post-edited), either especially created for the study, or taken from other MT/PE evaluation tasks. None of the above-mentioned studies has investigated post-editese on corpora produced by professional translators in their usual working conditions. Our project aims at filling this gap and exploring the possibilities for building and using comparable corpora for the study of post-editese.

We describe our methodology to build a first pilot corpus and the challenges we encountered during this process.

2. Corpus building

The material to create this pilot corpus was made available to us by the language service of the European Investment Bank (EIB) in the form of two translation memories (TMs) containing human translation (HT) and post-edited machine translation (PEMT) for the language direction English-to-French and for one specific domain.

We extracted and separated source, HT and PEMT to build two corpora (HT and PEMT), each comprising two subcorpora (source and target). Both corpora were pre-processed and cleaned to remove personal information, as well as non-alphabetical segments and duplicates (source-target pairs appearing more than once).

Corpora	HT	HT	PEMT	PEMT
Sub-Corpora	Source	Target	Source	Target
Nbr of sentences	3 440	3 440	1 327	1 327
Nbr of words	47 781	62 588	26 646	36 612
Mean sent. length (inc. punct)	16.08	21.32	22.66	31.68

Table 1: Length and average sentence length in words for each corpus after pre-processing.

Table 1 summarizes some of the corpora's statistics after pre-processing. We remark the important size difference between HT and PEMT corpora, which could constitute an issue since some metrics, such as TTR, can be sensitive to text length (Brezina, 2018). Another potentially problematic difference is the average source sentence length that is much smaller for HT than PEMT. For our first experiment, we tried to minimize these differences by using a random sampling as well as a sampling by source sentence length.

3. Pilot experiment

We performed a first pilot experiment to assess if the following commonly used metrics could be applied for the comparison of our HT and PEMT target corpora:

- type/token ratio or TTR (number of unique words/total number of words)
- ratio of content words (number of content words/total number of words)
- length ratio at sentence level (absolute difference between source and target length, normalised by the source length)

TTR and ratio of content words were computed for source and target to account for possible source variation between HT and PEMT. TTR was

significantly lower for PEMT than for HT in the target, but this was also the case in the source. The ratio of content words was higher in the target PEMT compared to HT, but we observed the opposite tendency in the sources. As we observed differences in the sources, TTR and ratio of content word variations in the target are difficult to interpret and conclusions on eventual post-editese features should be drawn carefully.

Finally, we computed the length ratio at sentence level for HT and PEMT. PEMT obtained a ratio significantly higher than HT. To ensure that this variation was not due to a difference in sentence lengths in the source corpora, we tested the correlation between the length ratio and the source length using Pearson's correlation test. No significant correlation was found.

4. Conclusion and perspectives

The creation of this pilot corpus for the study of post-editese in professional context has made us aware of the challenges of comparability while studying post-editese on comparable corpora. We saw that having equal amount of texts from the same company and domain is not necessarily sufficient to build a suitable comparable corpora and that comparability might also vary depending on the metrics used. Finally, our pilot experiment highlighted the importance of taking into account the potential variation in the source as well as the fact that common metrics for the study of post-editese might not all be pertinent when working with comparable corpora. For future work, we plan to increase the size of our corpus and to investigate different metrics (automatic, semi-manual and manual) for the study of the post-editese phenomenon with comparable corpora. Source corpora are also interesting to explore as they might reveal inherent differences between the source texts intended for PEMT or HT. The existence of such differences would show the importance of studying post-editese on corpora that reflect the actual use of PEMT in professional context.

References

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide.* Cambridge University Press.

Castilho, S. & Resende, N. (2022). Post-Editese in Literary Translations. *Information*, 13(2),66.

Castilho, S. Resende, N. & Mitkov, R. (2019). What Influences the Features of Post- editese? A Preliminary Study. *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*, 19-27.

Daems, J., De Clercq, O,, & Macken, L. (2017). Translationese and Post-editese: How comparable is comparable quality? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16, 89–103.

Toral, A. 2019. Post-editese: an Exacerbated Translationese. *Proceedings of MT Summit XVII*, 273-281.

••••