

The contribution of learner translation corpora to theory development in empirical translation studies: Overview and future prospects

Marie-Aude Lefer (UCLouvain, Belgium)

Corpus-based translation studies (CBTS) has considerably expanded its reach in recent years, with noteworthy advances on many fronts – theoretical, methodological, descriptive – and increased cross-fertilization and interdisciplinarity, both within translation studies and beyond (see De Sutter & Lefer 2020, Gaspari 2022, Kotze 2020 and 2022, Neumann et al. 2022, Vandevoorde et al. 2020). As shown by a recent research survey (Granger & Lefer 2022), CBTS relies strongly on parallel corpora, i.e. corpora containing source texts in a given language, aligned with their translations in another language. These corpora are typically made up of professional or expert translations into the translators' native language (Lefer 2020). Other translation types, such as student translations, non-professional translations, translations into the foreign/second language and machine translation post-editing, remain largely marginal in corpus research. The present talk deals with one of these underexplored corpus resources, namely learner translation corpora (LTC), and aims to show how LTC can contribute to theory development in empirical translation studies. LTC are parallel corpora containing authentic translations produced by translation students or foreign language learners in real-life tasks (e.g. in the translation classroom, as opposed to a laboratory setting; cf. Serbina & Neumann 2021), aligned with their source texts. In addition to representing novice, inexperienced translators, LTC differ from the parallel corpora traditionally used in CBTS in that many of them are *multiple* translation corpora, i.e. they include a number of translations of the same source text, often in large quantities. This unique design offers new opportunities for CBTS and, more generally, empirical translation studies.

The first LTC emerged two decades ago (e.g. PELCRA – Uzar & Walinski 2001, STA – Bowker & Bennison 2003) and were soon followed by similar initiatives (e.g. MISTiC – Castagnoli 2009, MeLLANGE – Castagnoli et al. 2011, NEST – Graedler 2013, UPF – Espunya 2014, RusLTC – Kutuzov & Kunilovskaya 2014, KOPE – Wurm 2016, CELTraC – Fictumova et al. 2017). In the talk, I will survey recent LTC-based studies in terms of their research objectives, objects of study, key findings and theoretical contributions. I will show that the main foci so far have been on computer-aided translation error analysis and translation quality evaluation (e.g. De Sutter et al. 2017, Espunya 2014, Kübler et al. 2022, Kunilovskaya & Lapshinova-Koltunski 2019, Vela et al. 2014), mostly with a view to informing translator training. Alongside this core applied-research strand, new types of LTC-based empirical investigation, with a somewhat stronger focus on theorization, have started to emerge in recent years. This is true, for instance, of the study of translation features such as explicitation, normalization and levelling-out, which had hitherto been approached mainly through the lens of expert translation in CBTS (see e.g. Castagnoli 2016, Kunilovskaya et al. 2018, Lapshinova-Koltunski 2022, Looock 2020, Redelinghuys 2016, Redelinghuys & Kruger 2015). I will also discuss LTC-based studies that have addressed key constructs such as translation competence acquisition (e.g. through longitudinal corpus studies; Wurm 2020) and translation variation (Castagnoli 2020).

The next part of my talk will be devoted to theory-oriented research based on the *Multilingual Student Translation* (MUST) corpus. MUST is an international LTC collection initiative that brings together 45 partner teams worldwide (Granger & Lefer 2020). The corpus currently comprises ca 400 source texts (ranging from 150 to 1,000 words in length) and 6,000 student translations produced by ca 3,000 students, with 19 languages represented. It includes numerous text types, both general (news and opinion articles, excerpts from novels, etc.) and specialized (financial reports, tourist guides, instruction manuals, contracts, etc.). The strengths of the MUST corpus include its rich standardized metadata relating to the source texts, translation tasks and learners (40+ metadata rubrics), and the Translation-oriented Annotation System (TAS) developed collaboratively within the MUST network to support research on translation quality across language pairs (Granger & Lefer 2021). In the talk, I will zoom in on two topics currently being investigated at UCLouvain in connection with the English-French pair, namely (1) simplification phenomena in student translations and their sources of explanation, tapping into the rich learner-related metadata collected in the MUST project and insights from contrastive linguistics (Penha-Marion et al. forthcoming) and (2) individual variation and translation invariance, relying on a combination of corpus-based and

corpus-driven linguistically informed analyses (pre-selected source-text features, such as figurative language, cognates, established/ad hoc noun sequences; target-text n-grams and POS-grams).

The talk will end with a look to the future. I will take stock of the LTC-derived theoretical advances made so far to sketch key avenues for LTC-based research, such as the link between translation variation, quality and literality, while also paying attention to the most acute challenges that lie ahead.

References

- Bowker, L. & Bennison, P. (2003). Student Translation Archive: Design, development and application. In F. Zanettin, S. Bernardini & D. Stewart (eds.) *Corpora in Translator Education*. London & New York: Routledge, 103-117.
- Castagnoli, S. (2020). Translation choices compared: Investigating variation in a learner translation corpus. In S. Granger & M.-A. Lefer (eds.) *Translating and Comparing Languages: Corpus-based Insights*. Corpora and Language in Use Proceedings 6. Louvain-la-Neuve: Presses universitaires de Louvain, 25-44.
- Castagnoli, S. (2016). Investigating trainee translators' contrastive pragmalinguistic competence: A corpus-based analysis of interclausal linkage in learner translations. *The Interpreter and Translator Trainer* 10(3), 343-363.
- Castagnoli, S. (2009). *Regularities and variations in learner translations: A corpus-based study of conjunctive explicitation*. PhD thesis, Pisa University.
- Castagnoli, S., Ciobanu, D., Kübler, N., Kunz, K. & Volanschi, A. (2011). Designing a Learner Translator Corpus for Training Purposes. In N. Kübler (ed.) *Corpora, Language, Teaching, and Resources: From Theory to Practice*. Bern: Peter Lang, 221-248.
- De Sutter, G., Cappelle, B., De Clercq, O., Looock, R., & Plevoets, K. (2017). Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translation. *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 16, 25-39.
- De Sutter, G. & Lefer, M.-A. (2020). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives* 28(1), 1-23.
- Espunya, A. (2014). The UPF learner translation corpus as a resource for translator training. *Language Resources and Evaluation* 48, 33-43.
- Fictumova, J., Obrušnik, A., & Stepankova, K. (2017). Teaching specialized translation error-tagged translation learner corpora. *Sendebär* 28, 209-241.
- Gaspari, F. (2022). Expanding the reach of corpus-based translation studies: The opportunities that lie ahead. In S. Granger & M.-A. Lefer (eds.) *Extending the Scope of Corpus-based Translation Studies*. London: Bloomsbury, 42-63.
- Graedler, A.-L. (2013). NEST—A corpus in the brooding box. *Studies in Variation, Contacts and Change in English*, 13.
- Granger, S. & Lefer, M.-A. (2022). Corpus-based translation and interpreting studies: A forward-looking review. In S. Granger & M.-A. Lefer (eds.) *Extending the Scope of Corpus-based Translation Studies*. London: Bloomsbury, 13-41.
- 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.

- Kotze, H. (2022). Translation as constrained communication: Principles, concepts and methods. In S. Granger & M.-A. Lefer (eds.) *Extending the Scope of Corpus-based Translation Studies*. London: Bloomsbury, 67-97.
- Kotze, H. (2020). Converging what and how to find out why: An outlook on empirical translation studies. In L. Vandevoorde, J. Daems & B. Defrancq (eds.) *New Empirical Perspectives on Translation and Interpreting*. Abingdon: Routledge, 333-370.
- Kübler, N., Mestivier, A., & Pecman, M. (2022). Using comparable corpora for translating and post-editing complex noun phrases in specialized texts: Insights from English-to-French specialized translation. In S. Granger & M.-A. Lefer (eds.) *Extending the Scope of Corpus-based Translation Studies*. London: Bloomsbury, 237-266.
- Kunilovskaya, M. & Lapshinova-Koltunski, E. (2019). Translationese features as indicators of quality in English-Russian human translation. In I. Temnikova, C. Orasan, G. Corpas Pastor & R. Mitkov (eds.) *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019)*, 47-56.
- Kunilovskaya, M., Morgoun, N., & Pariy, A. (2018). Learner vs. professional translations into Russian: Lexical profiles. *Translation and Interpreting* 10(1), 33-52.
- Kutuzov, A., & Kunilovskaya, M. (2014). Russian learner translator corpus: Design, research potential and applications. In P. Sojka, A. Horak, I. Kopecek & K. Palak (eds.) *Text, Speech and Dialogue. Lecture Notes in Computer Science*. Berlin: Springer, 315-323.
- Lapshinova-Koltunski, E. (2022). Detecting normalisation and shining-through in novice and professional translations. In S. Granger & M.-A. Lefer (eds.) *Extending the Scope of Corpus-based Translation Studies*. Bloomsbury, 182-206.
- Lefer, M.-A. (2020). Parallel corpora. In M. Paquot & S. Th. Gries (eds.) *A Practical Handbook of Corpus Linguistics*. Cham: Springer, 257-282.
- Loock, R. (2020). It's non-canonical word order that you should use! A corpus approach to avoiding standardized word order in translated French. In S. Granger & M.-A. Lefer (eds.) *Translating and Comparing Languages: Corpus-based Insights*. Corpora and Language in Use Proceedings 6. Louvain-la-Neuve: Presses universitaires de Louvain, 69-85.
- Neumann, S., Freiwald, J., & Heilmann, A. (2022). On the use of multiple methods in empirical translation studies: A combined corpus and experimental analysis of subject identifiability in English and German. In S. Granger & M.-A. Lefer (eds.) *Extending the Scope of Corpus-based Translation Studies*. London: Bloomsbury, 98-129.
- Penha-Marion, L. A. de S., Gilquin, G., & Lefer, M.-A. (forthcoming). The effect of directionality on lexico-syntactic simplification in French>English student translation. In H. Kotze & B. Van Rooy (eds.) *Constraints on Language Variation and Change in Complex Multilingual Contact Settings*. Amsterdam: John Benjamins.
- 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, 189-220.
- Redelinghuys, K. & Kruger, H. (2015). Using the features of translated language to investigate translation expertise: A corpus-based study. *International Journal of Corpus Linguistics* 20(3), 293-325.
- Serbina, T. & Neumann, S. (2021). Translation product and process data: A happy marriage or worlds apart? In S. L. Halverson & I. Marín García (eds.) *Contesting Epistemologies in Cognitive Translation and Interpreting Studies*. Abingdon: Routledge, 131-152.
- Uzar, R. & Walinski, J. (2001). Analysing the fluency of translators. *International Journal of Corpus Linguistics* 6, 155-166.
- Vandevoorde, L., Daems, J., & Defrancq, B. (eds.) (2020). *New Empirical Perspectives on Translation and Interpreting*. Abingdon: Routledge.

Vela, M., Schumann, A.-K., & Wurm, A. (2014). Beyond linguistic equivalence. An empirical study of translation evaluation in a translation learner corpus. In *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation*, 47-56.

Wurm, A. (2020). Translation quality in an error-annotated translation learner corpus. In S. Granger & M.-A. Lefer (eds.) *Translating and Comparing Languages: Corpus-based Insights*. Corpora and Language in Use Proceedings 6. Louvain-la-Neuve: Presses universitaires de Louvain, 141-162.

Wurm, A. (2016). Presentation of the KOPTE Corpus and Research Project.
https://www.academia.edu/24012369/Presentation_of_the_KOPTE_Corpus_and_Research_Project.