

Breaking new ground in contrastive and translation studies: Learner translation corpora to the fore

Marie-Aude Lefer
UCLouvain

marie-aude.lefer@uclouvain.be

Corpus-based contrastive and translation studies have considerably expanded their reach in recent years, with noteworthy advancement on many fronts – theoretical, methodological and descriptive – and increased interdisciplinarity (see e.g. De Sutter & Lefer 2020, Enghels et al. 2020, Granger & Lefer 2020a, Kotze forthcoming, Neumann et al. forthcoming, Vandevoorde et al. 2020). As confirmed by recent surveys of corpus-based contrastive linguistics (Hasselgård 2020) and translation studies (Granger & Lefer forthcoming), one thing that contrastive and translation studies have in common is their strong reliance on parallel corpora, i.e. corpora containing source texts in a given language, aligned with their translations in another language. These parallel corpora are typically made up of *professional* (or expert) translations into the translators' *native language*, with relatively few exceptions (Lefer 2020). In this talk, I aim to show how corpus-based contrastive and translation studies can break new ground by making use of *learner translation corpora* (LTC). LTC are a subtype of learner corpora (collections of texts produced by learners of a language; Gilquin 2020). More specifically, they are parallel corpora containing novice translations produced, for example, by translation students or foreign language learners. Being instances of both parallel corpora and learner corpora, LTC are situated at the interface of translation studies, contrastive linguistics and learner corpus research (Granger & Lefer 2020b).

Calls for combined analysis of bilingual comparable or parallel data and learner data first emanated from learner corpus research, with Granger's (1996) *Integrated Contrastive Model* (see Gilquin 2000/2001, Granger 2015, 2018 and Hasselgård & Ebeling 2018 for extensions of the model). The basic tenet of the model is that bilingual data can help shed light on transfer in learner language. The potential that lies in such combinations of corpus data

types was also acknowledged quite early on in both contrastive linguistics (Johansson 2007: 313) and translation studies (Chesterman 2007: 63). Even though the proposed approaches have so far been taken up only timidly in empirical investigations (e.g. Altenberg 2002, Behrens 2006, Ebeling & Hasselgård 2021, Gilquin 2008, Vanderbauwhede 2012), it is expected that they will soon be thriving, for instance under the impetus of the constrained language framework put forward by Kotze (2020, forthcoming) (see e.g. De Sutter & Lefer 2020, Ivaska et al. forthcoming). LTC represent yet another way of integrating learner data into contrastive and translation studies. To date, however, their use in empirical research has remained relatively marginal.

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, KOPTe - Wurm 2016, CELTraC - Fictumova et al. 2017). In this talk, I will sketch out the main features of the LTC compiled to date (language pairs, translation directionality, registers, etc.) and survey existing LTC-based research, in terms of research objectives, topics, corpus methodology and key findings. 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. 2018, Vela et al. 2014), mostly with a view to informing translation pedagogy and devising corpus-informed teaching materials. Alongside this core applied-research strand, new types of LTC-based empirical investigation have started to emerge in recent years, especially as regards the study of translation features, which had hitherto been approached mainly through the lens of expert translation (see e.g. Castagnoli 2016, Kunilovskaya et al. 2018, Lapshinova-Koltunski forthcoming, Loock 2020, Redelinghuys & Kruger 2015).

The next part of my talk will be devoted to the *Multilingual Student Translation* (MUST) project, an international LTC collection initiative which brings together more than 40 partner teams worldwide (Granger & Lefer 2020b). Among its many strengths, I will describe its shared source-text database, its rich standardized metadata related to the source texts, translation tasks and learners, and the Translation-oriented Annotation System (TAS) that is currently being developed collaboratively within the MUST network. I will also

present some MUST-based studies to illustrate the research potentialities of the corpus (e.g. Penha-Marion et al. forthcoming on translation directionality).

The talk will end with a few forward-looking remarks. In particular, I will discuss the various ways in which LTC can be combined with other corpus and observational data types to shed new light on cross-linguistic contrasts and translation. I will also argue that LTC data can help foster theoretical development, especially as regards the further elaboration of key constructs such as translation competence acquisition (e.g. through longitudinal studies; see Wurm 2020), translation expertise (e.g. Kajzer-Wietrzny 2020), default translation (Halverson 2019) and translation variation (e.g. Castagnoli 2020). Finally, I will sketch out the promising new opportunities for LTC-based research to remain relevant in today's world, such as the collection of LTC devoted to forms of interlingual mediation other than written translation (e.g. post-editing, subtitling, video game and web localization).

References

- Altenberg, B. (2002). Using bilingual corpus evidence in learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins, 37-54.
- Behrens, B. (2006). Language-based processing in advanced L2 production and translation: An exploratory study. In H. Byrnes, H. Weger-Guntharp & K. Sprang (eds.) *Educating for Advanced Foreign Language Capacities*. Georgetown University Press, 74-86.
- 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.
- Chesterman, A. (2007). Similarity analysis and the translation profile. *Belgian Journal of Linguistics* 21, 53-66.

- 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.
- Ebeling, S. & Hasselgård, H. (2021). The functions of n-grams in bilingual and learner corpora: An integrated contrastive approach. In S. Granger (ed.) *Perspectives on the L2 Phrasicon: The View from Learner Corpora*. Multilingual Matters.
- Enghels, R., Defrancq, B., & Jansegers, M. (eds.) (2020). *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*. Berlin: De Gruyter Mouton.
- Espunya, A. (2014). The UPF learner translation corpus as a resource for translator training. *Language Resources and Evaluation* 48, 33-43.
- Fictumova, J., Obrusnik, A., & Stepankova, K. (2017). Teaching specialized translation error-tagged translation learner corpora. *Sendebær* 28, 209-241.
- Gilquin, G. (2020). Learner corpora. In M. Paquot & S. Th. Gries (eds.) *A Practical Handbook of Corpus Linguistics*. Springer, 283-303.
- Gilquin, G. (2008). Combining contrastive and interlanguage analysis to apprehend transfer: detection, explanation, evaluation. In G. Gilquin, S. Papp & B. Díez-Bedmar (eds.) *Linking Contrastive and Learner Corpus Research*. Amsterdam & New York: Rodopi, 3-33.
- Gilquin, G. (2000/2001) The Integrated Contrastive Model: Spicing up your data. *Languages in Contrast* 3(1), 95-124.
- Graedler, A.-L. (2013). NEST—A corpus in the brooding box. *Studies in Variation, Contacts and Change in English*, 13.
- Granger, S. (2018). Tracking the third code: A cross-linguistic corpus-driven approach to metadiscursive markers. In A. Cermakova & M. Mahlberg (eds.) *The Corpus Linguistics Discourse*. Amsterdam & Philadelphia: John Benjamins, 185-204.
- Granger, S. (2015). Contrastive Interlanguage Analysis. A reappraisal. *International Journal of Learner Corpus Research* 1(1), 7-24.
- Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (eds.) *Languages in Contrast: Text-based Cross-Linguistic Studies. Lund Studies in English*. Lund: Lund University Press, 37-51.
- Granger, S. & Lefer, M.-A. (forthcoming). 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*. Bloomsbury.
- Granger, S. & Lefer, M.-A. (eds.) (2020a). *The Complementary Contribution of Comparable and Parallel Corpora to Crosslinguistic Studies*. Special issue of *Languages in Contrast*, 20(2).
- Granger, S. & Lefer, M.-A. (2020b). The *Multilingual Student Translation* corpus: a resource for translation teaching and research. *Language Resources and Evaluation* 54, 1183-1199.

- Halverson, S. L. (2019). 'Default' translation. A construct for cognitive translation and interpreting studies. *Translation, Cognition & Behavior* 2(2), 187-210.
- Hasselgård, H. (2020). Corpus-based contrastive studies: Beginnings, developments and directions. In S. Granger & M.-A. Lefer (eds.) *The Complementary Contribution of Comparable and Parallel Corpora to Crosslinguistic Studies*. Special issue of *Languages in Contrast* 20(2), 184-208 .
- Hasselgård, H. & Ebeling, S. (2018). At the interface between Contrastive Analysis and Learner Corpus Research: A parallel contrastive approach. *Nordic Journal of English Studies* 17(2),182-214.
- Ivaska, I., Ferraresi, A., & Bernardini, S. (forthcoming). Syntactic properties of constrained English: A corpus-driven approach. In S. Granger & M.-A. Lefer (eds.) *Extending the Scope of Corpus-based Translation Studies*. Bloomsbury.
- Johansson, S. (2007). *Seeing through Multilingual Corpora On the use of corpora in contrastive studies*. Amsterdam and Philadelphia: John Benjamins
- Kajzer-Wietrzny, M. (2019). Linking words in inter- and intralingual translation - Combining corpus linguistics and key-logging data. In L. Vandevoorde, J. Daems & B. Defrancq (eds.) *New Empirical Perspectives on Translation and Interpreting*. Routledge, 114-138.
- Kotze, H. (forthcoming). Translation as constrained communication: Principles, concepts and methods. In S. Granger & M.-A. Lefer (eds.) *Extending the Scope of Corpus-based Translation Studies*. Bloomsbury.
- 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-Volanschi, A., & Pecman, M. (2018). Teaching specialised translation through corpus linguistics: quality assessment and methodology evaluation by experimental approach. *Meta* 63(3), 806-824.
- 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. (forthcoming). 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.
- Lefer, M.-A. (2020). Parallel corpora. In M. Paquot & S. Th. Gries (eds.) *A Practical Handbook of Corpus Linguistics*. 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. (forthcoming). 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*. Bloomsbury.

- Penha-Marion, L. A. de S., Gilquin, G., & Lefer, M.-A. (forthcoming). Lexico-syntactic simplification in French><English student translation: The effect of translation directionality on bilingual language production. In H. Kotze & B. Van Rooy (eds.) *Constraints on Language Variation and Change in Complex Multilingual Contact Settings*. John Benjamins.
- 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.
- Uzar, R. & Walinski, J. (2001). Analysing the fluency of translators. *International Journal of Corpus Linguistics* 6, 155-166.
- Vanderbauwhede, G. (2012). The Integrated Contrastive Model evaluated: the French and Dutch demonstrative determiner in L1 and L2. *International Journal of Applied Linguistics* 22(3), 392-413.
- 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 KOPTÉ Corpus and Research Project. https://www.academia.edu/24012369/Presentation_of_the_KOPTÉ_Corpus_and_Research_Project.