WORKSHOP PROPOSAL - SLE ZÜRICH 2017

New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges

Convenors: *Renata Enghels, Marlies Jansegers, Sanne Tanghe*

Renata.Enghels@UGent.be
Marlies.Jansegers@UGent.be
Sanne.Tanghe@ugent.be

Ghent University
Linguistics Department- Spanish and Romance Linguistics
Blandijnberg 2
9000 Ghent
Belgium

http://research.flw.ugent.be/nl/renata.enghels
http://research.flw.ugent.be/nl/marlies.jansegers
http://research.flw.ugent.be/nl/sanne.tanghe

The practice of comparing languages has a long tradition characterized by a cyclic pattern of interest (Granger 2003; Schmied 2008). In the 1990s contrastive linguistics underwent a significant revival, which mainly originated from its meeting with corpus linguistics. This has led to a new wave of corpus-based contrastive studies. Still, until today there are two main challenges that have not yet been fully addressed: (1) an empirical assessment of the nature of the data which are commonly used in cross-linguistic studies (namely translation data vs. comparable data), and (2) the development of advanced methods and statistical techniques suitably adapted to the methodological challenges that are raised by contrastive research questions. This workshop aims to provide a forum through which these topics can be discussed.

The first challenge of contrastive linguistics relates to the variable nature of the empirical data it resorts to. Many contrastive linguists have turned to translation studies as a means of establishing cross-linguistic relationships (Granger, Lerot, Petch-Tyson 2003; Granger 2003; Johansson 2007). However, the use of parallel corpora as a source for contrastive linguistic research has not always gone undisputed. The most frequently cited disadvantages relate to (1) translation universals, i.e. "features which typically occur in translated text rather than original utterances" (Baker 1993: 243) and, (2) interference between the language of the source-text and the translated text (Johansson 1998; McEnery and Xiao 2008). However, contrary to these stated shortcomings, a commonly cited advantage of the use of parallel corpora relates to the *tertium comparationis*, i.e. a "common platform of comparison" (Connor & Moreno 2005: 157) against which differences can be described (James 1980; Johansson 2007; Granger 2010). The difficulty of establishing full comparability indeed constitutes one of the major stumbling blocks in the use of comparable corpora. Taking into account these limitations of both translation and comparable data, more recently, several linguists have argued in favor of a combination of the two, as complementary sources for cross-linguistic comparison (among others Viberg 2005; Altenberg and Granger 2002; Gilquin 2008; McEnery and Xiao 2008; Mortier and Degand 2009; Vanderschueren 2010; Enghels & Jansegers 2013; Jansegers 2017). However, up to the present, this combined corpus method has not yet been exploited to its full potential.

Moreover, in the last decade the analytical possibilities seem to have increased considerably as ever more multilingual data are made available. Contrastive linguistics not only benefits from the creation of huge web corpora (such as WebCorp and Sketch Engine), a growing number of new data types is becoming available, like subtitle corpora (e.g. Levshina, 2016) or the Wikipedia Parallel Titles Corpora. Besides the fact that the use of these resources is perhaps not yet widespread among linguists, the question of whether these different data can be applied to answer different contrastive research questions still remains to be answered.

A second challenge of contrastive linguistics relates to the methodological branch of corpus-based contrastive linguistics, which, according to Gast (2015: 5), "is still tender". Indeed, if a more advanced standard of methods and procedures is becoming common ground in monolingual studies (such as logistic and mixed-effects regression techniques, clustering analyses, cf. among other Gries 2013, Levshina 2015), the implementation of such techniques is still in its infancy in the field of contrastive linguistics. What is more, the more advanced methodological tools that are suited to study the multidimensional nature of linguistic

phenomena within one language, cannot be directly transferred to contrastive data without a thoughtful consideration, given the increased complexity of the latter. For instance, can language simply be taken as a response variable in regression models in order to compare different sensitivities to certain variables in the choice for certain linguistic expressions (as suggested for instance by Wiechmann 2011 in his study on relative clause constructions in English and German)?

This workshop aims at bringing together linguists working in different areas (synchronic, historical, and contrastive linguistics; translation or typological studies, etc.), and on different languages in order to reflect on the value and applicability of different kinds of empirical data for contrastive linguistics, and to contribute to methodological and theoretical advances in this domain. We particularly welcome submissions dealing with contrastive (case)studies making use of (more) rigorous empirically-based contrastive analyses (based on corpus data and/or experimental data) and/or making use of new data types, like subtitle corpora and web corpora. As such, we invite speakers to collectively discuss the methodological apparatus of Contrastive Linguistics, dealing with, but not limited to, the following questions:

1. How can we most efficiently make use of translation corpora for contrastive linguistics, while taking into account linguistic interferences and translation universals?
2. What (new) types of data are the most useful for what kind of contrastive questions?
3. Is it mandatory to complement translation data with comparable corpus data, or does this depend on the level of linguistic analysis (e.g. studies on lexical cognates vs. syntactic cognates vs. pragmatic phenomena, etc.)?
4. Which (advanced) statistical techniques are most suited to deal with the multidimensionality of contrastive research questions?
5. How can we go beyond a mere comparison of frequency tables between different comparable corpora?
6. How can we compare multifactoriality behind specific linguistic phenomena between two or more languages?

References

Altenberg, B., & Granger, S. (2002). Recent trends in cross-linguistic lexical studies. In B. Altenberg & S. Granger (Eds.), *Lexis in Contrast. Corpus-based approaches* (pp. 3-48). Amsterdam: John Benjamins.

Baker, M. (1993). Corpus linguistics and translation studies. Implications and applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology. In honour of John Sinclair* (pp. 233-250). Amsterdam: John Benjamins.

Connor, U. M. & Moreno, A. I. (2005). Tertium Comparationis: A vital component in contrastive research methodology. In P. Bruthiaux, D. Atkinson, W. G. Eggington, W. Grabe, & V. Ramanathan (Eds.), *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan* (pp. 153-164). Clevedon: Multilingual Matters.

Enghels, R. & Jansegers, M. (2013). On the cross-linguistic equivalence of *sentir(e)* in Romance languages: a contrastive study in semantics, *Linguistics* 51(5), 957-991.

Gast, V. (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.

Gilquin, Gaëtanelle (2008). Causative Make and Faire: A Case of Mismatch. In M. Gómez González, J. L. Mackenzie & E. M. González Álvarez (Eds.), *Current Trends in Contrastive Linguistics: Functional and Cognitive Perspectives* (pp.177-201). Amsterdam: John Benjamins.

Granger, S. (2003). The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies? In S. Granger, J. Lerot & S. Petch-Tyson (Eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* (pp. 17-29). Amsterdam: Rodopi.

Granger, S. (2010). Comparable and translation corpora in cross-linguistic research. Design, analysis and applications, *Journal of Shanghai Jiaotong University*, 2, 14-21.

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

Gries, S. T. (2013). *Statistics for linguistics with R. A practical introduction*. Berlin: Mouton De Gruyter.

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

Jansegers, M. (2017). *Hacia un enfoque múltiple de la polisemia. Un estudio empírico del verbo multimodal 'sentir' desde una perspectiva sincrónica y diacrónica.* Berlin: Mouton De Gruyter.

Johansson, S. (1998). On the role of corpora in cross-linguistic research. In S. Johansson & S. Oksefjell (Eds.), *Corpora and Cross-linguistic Research. Theory, Method, and Case Studies* (pp. 3-24). Amsterdam: Rodopi.

Johansson, S. (2007). *Seeing through Multilingual Corpora. On the use of corpora in contrastive studies*. Amsterdam: John Benjamins.

Levshina, N. (2015). *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Levshina, N. (2016). Verbs of letting in Germanic and Romance languages: A quantitative investigation based on a parallel corpus of film subtitles. *Languages in Contrast* 16(1): 84-117.

McEnery, T. & Xiao, R. (2008). Parallel and comparable corpora: what is happening? In G. Anderman & M. Rogers (Eds.), *Incorporating corpora. The Linguist and the Translator* (pp. 18-31). Clevedon: Multilingual Matters Ltd.

Mortier, L., & Degand, L. (2009). Adversative discourse markers in contrast. The need for a combined corpus approach, *International Journal of Corpus Linguistics*, 14(3), 301-329.

Schmied, J. (2008). Contrastive Corpus Studies. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook* (pp. 1140-1159). Berlin: Mouton De Gruyter.

Vanderschueren, Clara (2010). The use of translations in linguistic argumentation. A case study on Spanish and Portuguese subordinate clauses introduced by *para*, *Languages in Contrast* 10(1), 76-101.

Viberg, Å. (2005). The lexical typological profile of Swedish mental verbs, *Languages in Contrast* 5(1), 121-157.

Wiechmann, D. (2011). Exploring probabilistic differences between genetically related languages, *Languages in* Contrast, 11(2), 193-215.