Genre variation in translation corpora:
Evaluative prefixation in TED talks and parliamentary debates

Several scholars have recently emphasized the need to explore the variation across registers, genres and domains in corpus-based contrastive and translation studies. Johansson (2007: 304), for example, argues that "it is desirable to extend contrastive studies by taking into account the variation across registers within languages." Similarly, Olohan (2004: 191) states that "we need to recognize that a substantial part of translation activity in the world involves non-literary texts. If we are to study the activity of translation and translation process generally, we need to delve deeper into those genres and texts too. This will have the consequence that we will be able to make more cross-genre comparisons and study the extent to which features of translation may be influenced by genre, text type, etc."

However, in contrastive and translation studies based on translation corpora, the emphasis has mainly been on literary genres (such as novels) and relatively few studies have examined cross-genre variation (notable exceptions are Altenberg, 2004/2005; Cosme, 2006; Neumann, 2008; De Sutter et al., 2012). The present paper aims to contribute to this emerging body of corpus-based research and addresses the issue of genre variation in the field of contrastive word-formation. This will be done by examining English and French evaluative prefixation (e.g. En. *extra-high*, *hyper-innovative*, *mini-victory*, *outperform*) across two genres: talks and parliamentary debates. In fact, such variationist approaches are essential in word-formation studies, as word-formation is highly register-, genre- and domain-sensitive. As shown by Plag et al. (1999), for example, the use and productivity of English suffixes greatly vary across speech and writing (see Grabar et al., 2006; Chmielik & Grabar, 2011; Lefer, 2012 for similar findings about French).

Evaluative morphology has been extensively discussed in morphological typology (cf. Stump, 1993; Bauer, 1997; Grandi & Montermini, 2005; Körtvélyessy & Stekauer, 2011). Corpus-based contrastive descriptions of evaluative morphology, however, are still sorely lacking for many language pairs (see Andor, 2005), including English and French. Available descriptions of evaluative prefixation, which is the main focus of this paper, show that prefixes can be classified along the following two dimensions (see Wierzbicka, 1991; Grandi, 2002; Fradin & Montermini, 2009; the sub-categories are taken from Cartoni, 2008: 287-291):

Quantity dimension with a maximum/minimum axis (so-called 'measurativity') and the two semantic values 'BIG' and 'SMALL':
- 'BIG': increase, abundance (*hyper-, macro-, maxi-, mega-, super-*)
- 'SMALL': decrease, attenuation, approximation (*micro-, mini-; semi-, quasi-, pseudo-*)

Quality dimension with a positive/negative axis (so-called 'appreciativity') and the two semantic values 'GOOD' and 'BAD':
- 'GOOD': excess (excessive degree), superiority (higher rank) (*extra-, hyper-, mega-, out-, over-, super-, ultra-*)
- 'BAD': lack, inferiority (lower rank) (*hypo-, under-, sub-*)

The present study relies on English$_{SL}$-French$_{TL}$ bilingual data extracted from two translation corpora, which together total 20+ million words: the Europarl corpus of parliamentary debates (Koehn, 2005; Cartoni & Meyer, 2012) and WIT^3 (*Web Inventory of Transcribed and Translated Talks*, TED talks; Cettolo et al., 2012). Both corpora have been aligned at word-level by relying on a tailor-made alignment programme combined with Giza++ (Och & Ney, 2000) so as to facilitate the automatic extraction of translation equivalent pairs.

Our talk will focus on the impact of genre variation on (1) evaluative prefixation in English source texts (in terms of type/token frequencies, word categories, base words, semantics: qualitative vs quantitative evaluation) and (2) the translation equivalents used in French target texts. The French equivalents will be classified into the following two categories: incongruent translations (single words, e.g. *out-compete* - *supplanter*; and periphrastic translations, e.g. *over-bureaucratic* - *excessivement bureaucratique*) and congruent translations (evaluative prefixation, e.g. *semi-wild* - *semi-sauvage*,

underfunded - sous-financé). The corpus-based analysis will make it possible to measure the impact of genre on the word-formation features of both source and target texts and determine to what extent translation strategies are genre-dependent.

References

Altenberg B. (2004/2005). The generic person in English and Swedish. A contrastive study of *one* and *man*. *Languages in Contrast*, 5(1): 93-120.

Andor J. (2005). A Lexical Semantic-Pragmatic Analysis of the Meaning Potentials of Amplifying Prefixes in English and Hungarian. A Corpus-based Case Study of Near Synonymy. In *Proceedings from the Corpus Linguistics Conference Series*, 1(1). Available online: http://www.corpus.bham.ac.uk/PCLC/

Bauer L. (1997). Evaluative Morphology: In Search of Universals. *Studies in Language*, 21(3): 533-575.

Cartoni B. (2008). *De l'incomplétude lexicale en traduction automatique : vers une approche morphosémantique multilingue*. PhD thesis. Université de Genève: Genève.

Cartoni B. & T. Meyer (2012). Extracting Directional and Comparable Corpora from a Multilingual Corpus for Translation Studies. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (LREC), May 2012, Istanbul, Turkey.

Cettolo M., C. Girardi & M. Federico (2012). WIT3: Web Inventory of Transcribed and Translated Talks. In *Proceedings of EAMT*, Trento, Italy, 261-268.

Chmielik J. & N. Grabar (2011). Détection de la spécialisation scientifique et technique des documents biomédicaux grâce aux informations morphologiques. *TAL*, 52(2): 151-179.

Cosme C. (2006). Clause combining across languages. A corpus-based study of English-French translation shifts. *Languages in Contrast*, 6(1): 71-108.

De Sutter G., I. Delaere & K. Plevoets (2012). Lexical lectometry in corpus-based translation studies. Combining profile-based correspondence analysis and logistic regression modelling. In M. Oakes & J. Meng (eds) *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam: Benjamins.

Fradin B. & F. Montermini (2009). La morphologie évaluative. In B. Fradin, F. Kerleroux & M. Plénat (eds) *Aperçus de morphologie du français*. Saint Denis: PUV, 231-266.

Grabar N., D. Tribout, G. Dal, B. Fradin, N. Hathout, S. Lignon, F. Namer, C. Plancq, F. Yvon & P. Zweigenbaum (2006). Productivité quantitative des suffixations par *-ité* et *-able* dans un corpus journalistique moderne. In *Proceedings of TALN 2006 conference*, Leuven, Belgium.

Grandi N. (2002). *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*. Milan: FrancoAngeli.

Grandi N. & F. Montermini (2005). Prefix-Suffix Neutrality In Evaluative Morphology. In G. Booij, E. Guevara, A. Ralli, S. Sgroi & S. Scalise (eds) *Morphology and Linguistic Typology*. *Online Proceedings of the Fourth Mediterranean Morphology Meeting* (MMM4), September 2003, Catania, University of Bologna, 2005, 143-156.

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

Koehn P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of MT Summit X*, 79-86.

Körtvélyessy L. & P. Stekauer (eds) (2011). *Diminutives and Augmentatives in the Languages of the World*. *Lexis*, 6.

Lefer M.-A. (2012). La préfixation française à travers les genres et les domaines : étude de corpus. In *Proceedings of the 'Congrès Mondial de Linguistique Française'* (CMLF 2012), 1325-1349.

Neumann S. (2008). *Contrastive Register Variation. A quantitative approach to the comparison of English and German*. Habilitationsschrift. Universität des Saarlandes.

Och F.J. & H. Ney (2000). Improved Statistical Alignment Models. In *Proceedings of ACL*, 440-447.

Olohan, M. (2004). *Introducing Corpora in Translation Studies*. London: Routledge.

Plag I., C. Dalton-Puffer & H. Baayen (1999). Morphological productivity across speech and writing. *English Language and Linguistics*, 3(2): 209-228.

Stump G. T. (1993). How peculiar is evaluative morphology? *Journal of Linguistics*, 29: 1-36.

Wierzbicka A. (1991). *Cross-cultural pragmatics: The semantics of human interaction*. Berlin: Mouton de Gruyter.