# Different translators, different translation solutions? Analyzing n-grams in a French-to-Dutch multiple-translation corpus – and what it reveals about expertise, priming and cognitive routinisation

**Gert De Sutter (Ghent University), Marie-Aude Lefer (UCLouvain) & Koen Plevoets (Ghent University)**

KEYWORDS: translation variance, expertise, priming, cognitive routinization, controlled corpus design

In its 30 years of existence, corpus-based translation studies has resorted to a wide range of corpora, including comparable corpora (containing translated and non-translated texts in the same language), parallel corpora (containing source texts in language A and target texts in language B), intermodal corpora (containing texts in different translation modalities) and learner corpora (containing translations produced by student translators) (cf. Lefer, 2020). Recently, scholars have started to use so-called multiple-translation (parallel) corpora, consisting of different source texts that have each been translated multiple times in the same target language by different translators. This type of corpus enables a.o. an analysis of frequent translation solutions as well as a comparison of different translations for a given source segment (cf. Malmkjær, 1998, Castagnoli, 2020, 2023, Prinzie, 2022).

The present paper adds to this line of research by investigating how often specific n-grams (3 ≤ n ≤ 6) are translated identically, while at the same time relating these n-gram translation frequencies to the level of expertise (student vs. professional) and two well-known cognitive constraints, viz. structural priming (i.e. the re-use of a specific structure in subsequent discourse; Hartsuiker et al., 2016) and cognitive routinization (i.e. the degree of entrenchment of a specific structure in memory; Diessel, 2019). By doing so, we aim to find out (i) whether students show less variation in their translation solutions than professionals, (ii) whether structurally similar source text structures increase the likelihood of translators choosing an identical solution (i.e. structural priming) and (iii) whether the entrenchment (or cognitive routinization) of the chosen solution in the target language and the entrenchment of the cross-linguistic correspondence between source and target structures increase the likelihood of translators choosing the same solution.

To achieve these goals, we created a controlled parallel corpus of French-to Dutch news translations, which were produced by 7 undergraduate students of translation and 7 professionals (it is a subpart of the corpus used in De Sutter, Lefer and Vanroy, 2023). All translators were asked to translate 5 source texts from French to Dutch about a general journalistic topic for a non-specialised audience in the Dutch language area. They were not allowed to use machine translation or translation memories, or to ask another translator to revise their work, but they were free to use other (digital and non-digital) resources (e.g. dictionaries, glossaries, terminological databases). From this controlled parallel corpus we extracted from the Dutch translations all 3- to 6-grams with a minimal frequency of 3 using a custom-made Python script. After manual validation, the n-grams and their frequencies were exported to Excel for further annotation: we coded each of the n-grams for expertise (professional vs. student), for structural priming from the source text (primed vs. not primed; if the POS of each token in the source and target n-gram is identical, this instance was annotated as ‘primed’, otherwise: not primed), and for cognitive routinisation (numerical value; we used the independent Dutch Web 2020 and French Web 2020 corpus as well as the Dutch Parallel Corpus 2.0 (all available on Sketch Engine; Kilgarriff et al., 2014) to extract the relative frequencies of each of the n-grams in non-translated Dutch as well as the relative frequencies of the cross-linguistic correspondence between the source and target n-gram). This dataset was exported to R for statistical analysis (R Core Team 2016). First, we performed a Principal Component Analysis (PCA) to explore the variation in the full dataset (i.e. all ngrams with a minimal frequency of 3 in the controlled corpus) as well for each ngram width separately (i.e. a separate analysis for the 3-grams, one for the 4-grams and so on). The results of this PCA indicate that by and large the dispersion of the datapoints are dependent on text topic on the one hand and on expertise on the other hand. More particularly, students’ texts show more dispersion than professionals’ texts, indicating that students use a more diverse set of translation solutions given a specific source-text stimulus. Subsequently, we tested the hypothesis that the frequency of a specific translation solution (with a higher frequency indicating a larger homogeneity among the translators) is determined by expertise, structural priming and cognitive routinization, using a mixed-effects regression analysis. In our talk, we will analyze the results of this latter analysis against the background of Halverson’s (2017) cognitive model of gravitational pull and magnetism.

**References**

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*. *Selected Proceedings of the Fifth Using Corpora in Contrastive and Translation Studies Conference. Corpora and Language in Use Proceedings 6* (pp. 25–44). Presses Universitaires de Louvain.

Castagnoli, S. (2023). Exploring variation in student translation. *International Journal of Learner Corpus Research,* *9*(1), 97­–125.

Diessel, H. (2019). *The grammar network*. Cambridge University Press.

Halverson, S. (2017). Gravitational pull in translation. Testing a revised model. In G. De Sutter, M. Lefer, & I. Delaere (Eds.), *Empirical Translation Studies: New Methodological and Theoretical Traditions* (pp. 9–46). De Gruyter.

Hartsuiker, R. J., Beerts, S., Loncke, M., Desmet, T., & Bernolet, S. (2016). Cross-linguistic structural priming in multilinguals: Further evidence for shared syntax. *Journal of Memory and Language, 90*, 14–30.

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*, 7–36.

Lefer, M.-A. (2020). Parallel corpora. In M. Paquot, & S. Th. Gries, *Practical Handbook of Corpus Linguistics* (pp. 257–282). Springer.

Malmkjær, K. (1998). Love thy Neighbour: Will Parallel Corpora Endear Linguists to Translators? *Meta: Translators' Journal,* *43*(4), 534–541.

Prinzie, T. (2022). Variability in a multiple translation corpus as evidence for cognitive processes. Paper delivered at the *CogLing Days (Belgium Netherlands Cognitive Linguistics Association): Language, Discourse and Cognition*.

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.