

Reference

Granger, S. & Lefer, M.-A. (2020). Introduction: A two-pronged approach to corpus-based crosslinguistic studies. 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): 167-183.

Introduction

A two-pronged approach to corpus-based crosslinguistic studies

Sylviane Granger and Marie-Aude Lefer
Université catholique de Louvain (Belgium)

1. Corpora for crosslinguistic research

This special issue focuses on the combined use of parallel and comparable corpora, a methodology that was advocated by Johansson as early as 1998 but whose value for crosslinguistic research is only now beginning to be truly appreciated. As the terms ‘parallel’ and ‘comparable’ have received several definitions in the literature, it is essential to provide precise definitions of the two terms. Comparable corpora are corpora that are compiled according to the same sampling method and are therefore matched in terms of register, domain, period, etc. In corpus-based translation studies (CBTS), they are typically monolingual. For Baker (1995: 234), for example, “[c]omparable corpora consist of two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages”. In corpus-based contrastive linguistics (CBCL), on the other hand, they are typically multilingual, i.e. they “contain original texts in two or more languages matched by criteria such as genre, time of publication, etc.” (Johansson, 2007: 9). In this volume, the term ‘comparable corpus’ is used in the latter meaning. There appears to be more unanimity today as regards the term ‘parallel corpus’ (also called ‘translation corpus’), which generally refers to a collection of texts in one language and their translations into one or more languages. A parallel corpus is unidirectional if it only contains translations from one language into another. It is bidirectional if it contains translations in both directions (from language A to language B and vice versa).

Comparable corpora are easier to compile than parallel corpora. As most texts are available in electronic format today, it is possible to collect them in vast quantities in a wide range of registers. The main issue, however, is comparability. Kenning (2010: 490), for example, points to the difficulty of ensuring that newspaper corpora are truly comparable: “one finds a huge number of possible variations in terms of types and degrees of similarity: the type of newspaper, its standing, the genres to be selected and their function (...) not to mention factors such as number and size of articles and subject domain”. Parallel corpora, on the other hand, are much harder to compile, for several reasons: not all texts are translated or they may be translated from one language to another but much less frequently in the other direction. As a result, parallel corpora are usually quite small by today’s standards. They also tend to represent a narrow range of registers (mainly fiction and news), although there are some exceptions such as the CroCro corpus (Hansen-Schirra *et al.*, 2012) and the Dutch

Parallel Corpus (DPC) (Macken *et al.*, 2011). These smaller collections are directional, i.e. the source and the target languages are clearly identified (Cartoni and Meyer, 2012). The larger collections (e.g. institutional documents, subtitles, technical manuals and web-crawled texts) are often non-directional, which is not felt to be a major hindrance for natural language processing (NLP) applications but seriously reduces their usefulness for CBCL and CBTS.

While the two types of corpus are used in CBCL and, to a lesser extent, CBTS, they serve to answer different research questions and operate within quite different theoretical and methodological frameworks. In the following two sections we look at the role they play in each discipline before giving an overview of the volume in the last section.

2. The view from CBCL

There is some controversy in CBCL as to which of the two types of corpus – comparable or parallel – is *the* ideal resource for contrastive studies. All contrastive linguists agree that comparable corpora are an essential resource as they contain original texts in two or more languages, which makes them highly reliable for contrastive linguistic analysis. For Johansson (2007: 10), for example, “[c]omparable corpora of original texts have the advantage that they represent ordinary language use in each language and should allow safe conclusions on similarities and differences between the languages compared”. Lauridsen (1996: 67; 70) even considers that comparable corpora are the only reliable corpus type for contrastive linguistic analysis. Size is often an issue, however, the comparable corpus used being too small to investigate relatively infrequent linguistic phenomena. To solve the problem of data sparseness, Bossuyt and Leuschner (2020) suggest combining small comparable multilingual corpora with larger, non-comparable monolingual corpora for each of the languages investigated.

Although comparisons based on comparable corpora are ‘safe’, they are not necessarily easy to make. As pointed out by Johansson (2007: 10), “[t]he most difficult problem is knowing what to compare, i.e. relating forms which have similar meanings and pragmatic functions in the languages compared”. Identifying “*prima facie* equivalents” (Tognini-Bonelli, 2001: 134–135) is difficult on the sole basis of comparable corpora. In the case of lexis, the examples discussed by Tognini-Bonelli show that researchers often opt for the intuitively most obvious candidates, in particular cognates (e.g. English *in case of* and Italian *in caso di* or English *real* and Italian *reale*). But in many cases, it is not individual forms or phrases that are compared but whole categories – grammatical, semantic or discursive. To give but a few examples, Xiao *et al.* (2006) carry out a systematic analysis of passive constructions in British English and Mandarin Chinese speech and writing. Rozumko (2016) compares the repertoires of English and Polish adverbs of certainty based on the British National Corpus and the National Corpus of Polish. Pérez Blanco (2016) investigates negative evaluative adjectives in English and Spanish newspaper opinion discourse. Hasselgård (2014) focuses on discourse-structuring functions of initial adverbial adjuncts in English and Norwegian news and fiction. But even when whole categories are compared rather than individual forms, it remains unsure whether all the relevant comparable items have been included.

When it comes to equivalence, nothing beats parallel corpora. A parallel corpus gives access to what James (1980: 178) considers to be the best type of *tertium comparationis* for contrastive analysis, i.e. translation equivalence. For Salkie (2007[1997]: 344), although comparable corpora are favoured by some researchers, “parallel corpora of translated texts are the best source of data in this area”. For Mauranen (1999: 182), they are “indispensable for contrastive language study. Not only do they provide us with a good deal of data to work on, but they also expose the variety, indeterminacy and fuzziness of real bilingual language use, and keep us from setting up excessively tidy and objective contrasts between languages”. They constitute a powerful heuristic for the identification of correspondences between languages in the form of translation paradigms, i.e. “the set of forms in the target text which are found to correspond to particular words or constructions in the source text; or the other way around: the set of forms in the source text which are found to correspond to particular words or constructions in the target text” (Johansson, 2007: 23). Bidirectional parallel corpora

make it possible to compute what Altenberg (1999: 254) has called “the ‘mutual correspondence’ (MC) of categories and items in the two languages, i.e. the frequency with which different (grammatical, semantic and lexical) expressions are translated into each other”. This measure has revealed that words which at first sight appeared to be equivalent in fact displayed varying degrees of correspondence which could be greater in one direction than in the other, leading to “translation bias” (*ibid.*: 258). For example, Altenberg’s (1999: 258) study of adverbial connectors in English and Swedish showed that “Swedish *emellertid* is more often translated into English *however* (81%) than the other way round (47%)”. One of the strengths of parallel corpora lies in the identification of less straightforward equivalents, often belonging to different grammatical categories. Aijmer (1999), for example, shows that in almost 50% of cases the English modal *may* is not translated by a modal verb in Swedish but by an adverb (*kanske* ‘perhaps’), while modals are the default choice in Swedish-to-English translations. Ebeling and Oksefjell Ebeling (2015) have introduced a variant of the MC index called the reverse mutual correspondence (rMC), which complements Altenberg’s original index. Based on the translations and their sources, the new measure exploits the bidirectional structure of the corpus from a different angle by calculating the number of times the two items compared have each other as source. One important limitation of the use of parallel corpora in CBCL is that “it is excessively biased in favor of written language. Findings are therefore only poorly generalizable” (Enghels *et al.*, 2020: 9).

In spite of the undeniable advantage they present in terms of equivalence, parallel corpora have been the subject of a great deal of criticism among contrastive linguists. Lauridsen (1996: 67), for example, views them as unreliable and advises against their use: “one should refrain from using translation corpora unless the purpose of the linguistic analysis is either to evaluate the translation process or to criticize the translation product on the basis of a given translation theory”. Likewise, Gellerstam (1996: 53) observes that “[i]n cross-linguistic corpus studies, translations are sometimes frowned upon. The reason is obvious: if the target language is influenced by the structure of the source language, which is a well-known phenomenon (...), your study will not be based on two independent languages”. For McEnery *et al.* (2006: 93), “translated language is at best an unrepresentative special variant of the target language. If this special variant is confused with the target L1 language and serves alone as the basis for contrastive studies, the results are clearly misleading”. More recently, Aijmer and Lewis (2017: 3), while acknowledging the benefits of parallel corpora, warn that “[t]ranslations should be used with caution, however. The disadvantages of using parallel corpora is that they may suffer from ‘translationese’ (...), source-text influence, the translator’s fingerprints, and uneven translation quality”. Enghels *et al.* (2020: 13) also warn that “parallel (i.e. translated) data are to be handled with care”.

The respective strengths and weaknesses of comparable and parallel corpora led Johansson (1998) to put forward a corpus design which integrates the two. The main advantage of this design, which Egan (2019) refers to as the “4-text solution”, is that “it is not necessary to choose between parallel corpora and comparable corpora. Both can be combined within the same overall framework” (Johansson, 2011: 127). The English Norwegian Parallel Corpus (ENPC), initiated by Johansson in the 1990s, was the first corpus to be based on this model. It is a bidirectional translation corpus which contains English texts translated into Norwegian and Norwegian texts translated into English. As a result, “a comparable corpus and a translation corpus are combined within the same overall framework, and each type can be used to control and supplement the other” (Johansson, 2007: 11). Since then, the ENPC has served as a model for several other similarly designed corpus collections, such as the English-Swedish Parallel Corpus (Altenberg and Aijmer, 2000), the English-Portuguese COMPARA corpus (Frankenberg-Garcia and Santos, 2003) and the English-German CroCo corpus (Hansen-Schirra *et al.*, 2012).

Comparable and parallel corpora can be combined in two different ways. The approach which, as pointed out by Johansson (2007: 33), “is more in agreement with the traditional procedure of contrastive analysis (description, juxtaposition, and comparison), is to start with an analysis based on comparable data and then study translation correspondences”. The other approach is to take translations as the basis of comparison and then check the results against comparable corpus data. Each approach has its advantages and limitations, and the choice

depends on the types of linguistic phenomenon investigated and the research questions addressed.

Over the years Johansson's model has become increasingly popular in CBCL. For De Baets *et al.* (2020: 104), it is "a crucial procedure, and an inevitable step for contrastive-linguistic researchers who want to use data from parallel corpora for a contrastive study". Even corpus linguists who raise concerns about parallel corpora are keen to highlight their usefulness when combined with comparable corpora. For McEnery *et al.* (2006: 94), for example, parallel corpora are "a useful starting point" for further cross-linguistic analysis based on comparable corpora. In particular, the model allows for quantitative and qualitative comparison between original and translated language, thereby making it possible "to control for translation effects" (Johansson, 2007: 10). For example, a comparison of the verbs *love* and *hate* and their Norwegian standard equivalents *elske* and *hate* shows that *elske* and *hate* are used more frequently in texts translated from English than in original Norwegian texts (*ibid.*: 96). Johansson attributes this overuse to translators' insufficient awareness of an important difference between the verbs in the two languages: the Norwegian verbs *elske* and *hate* express a strong feeling and are most commonly used with a personal object, while the English verbs *love* and *hate* are also used in a weakened sense with non-personal objects (*love soap operas, hate travel*). Although the variety of other translations used by translators shows that they are generally aware of the difference, "[t]here is nevertheless a tendency to equate the English and Norwegian verbs and treat them as standard equivalents regardless of context" (*ibid.*: 102).

It is important to note that in CBCL translated language is only looked at against original, non-translated language to control for source text influence. So the comparison clearly has an ST orientation. This emerges very strongly from the following quote by Johansson (2007: 33): "More general features characteristic of translated texts (of the type discussed by Baker 1993 and others) will not be specifically discussed in this book. Where translation effects are mentioned, the reference is usually to features induced by the source text". In this perspective the term "translation-oriented CL" (Ramón García, 2002: 400; Ibrahim, 2015: 381) is rather misleading, as the main aim is to check the validity of systemic differences between languages rather than to investigate translation effects. However, things are changing and the CBCL/CBTS perspectives are gradually coming together. This is shown, for example, by Xiao's (2011) study of word clusters and reformulation markers in Chinese and English, which includes a discussion of the findings in the light of the explicitation hypothesis.

3. The view from CBTS

The combined use of comparable and parallel corpora has also been advocated in CBTS. Bernardini (2011: 12), for example, suggests that "a tripartite corpus structure (source texts in language A, target texts in language B and comparable/reference originals in language B) is the minimal setup to start to shed light on the distinctive features of translated language". However, the reasons for using parallel corpus data and the preferred type of comparable corpus are both different in CBTS and CBCL.

First, parallel corpora are used for very different purposes in the two fields. In CBTS, the main objective is not to identify translation equivalents or cross-linguistic correspondences. Rather, parallel corpora are mostly used with a view to investigating the main translation techniques (also called procedures, solutions or strategies) used to render a given item, pattern or phenomenon, whether lexical, syntactic, discursive or pragmatic. The scope of such translation research based on parallel corpora ranges from culture-specific items and other phenomena that are difficult to translate because they have no obvious equivalent in the target language (e.g. Marco, 2019, on food-related culturemes in literary translation), through translation features such as the explicitation and normalization of given source text items or structures (e.g. Denturck, 2012, on causal connectives in French-Dutch translation and Frankenberg-Garcia, 2019, on splitting and joining sentences in Portuguese-English translation), to translator style, i.e. stylistic idiosyncrasies in translation (e.g. Winters, 2013, on modal particles and Mastropierro, 2018, on key clusters). Another striking

difference between contrastive and translation studies based on parallel corpora is that CBCL typically relies on parallel corpora made up of written translations (especially fiction and non-fiction), while CBTS examines other forms of translation as well (cf. Enghels *et al.*, 2020). These include audio-visual translation such as subtitling and dubbing (e.g. Pavesi, 2013) and simultaneous interpreting (e.g. Magnifico and Defrancq, 2020). The integration of these types of parallel corpus is only starting to emerge in CBCL (see Levshina, 2016; Defrancq and Collard, 2020; Lefer and De Clerck, forthcoming).

Second, the type of comparable corpus most commonly used in CBTS is different from that used in CBCL. CBCL traditionally relies on multilingual comparable corpora, i.e. corpora that contain original texts in two or more languages (see Section 2). This type of comparable corpus is not (yet) frequently used in CBTS, which favours monolingual comparable corpora, i.e. corpora made up of translated and non-translated (original) texts in a given language. Accordingly, corpus triangulation in CBTS typically consists in supplementing monolingual comparable corpora with parallel corpora (or the other way around). Following Baker (1993, 1995), monolingual comparable corpora are used in CBTS to study the distinguishing features of translated language, such as explicitation, simplification and normalization, when compared with original language production or other forms of interlingual mediation (e.g. L2 writing) (see e.g. Kotze, 2019). Parallel corpora used in combination with monolingual comparable corpora make it possible to control for source-language (or source-text) influence, thereby helping to disentangle translation-inherent features, which are the main foci of CBTS, from source-language or -text influence (see e.g. Dai and Xiao, 2011, and Kenny and Sathachai, 2018, on the passive in English-Chinese and English-Thai translation respectively). In fact, the role of source-language influence in CBTS was explicitly set aside in Baker's (1993) research agenda for CBTS, which advocated recourse to monolingual comparable corpora for the purpose of investigating translation. As pointed out by Bernardini (2011: 2), at the time parallel corpora "appeared to perpetuate a source-oriented, prescriptive view of translated texts as almost by definition inferior to their source texts". As a result, source-language influence was often neglected in the early CBTS research dominated by the monolingual comparable approach proposed by Baker. It has, however, been reappraised in recent years, with translation scholars convincingly showing its crucial role in shaping translation products. Becher (2011), for example, shows that the addition and omission of connectives in German-English translation, usually taken in CBTS as instances of explicitation and implicitation, can be attributed to well-known German-English contrasts in terms of syntax, lexis, and communicative norms (see also Teich, 2003; Delaere and De Sutter, 2017).

Another way of taking cross-linguistic contrasts on board in translation research is to combine multilingual comparable corpora, monolingual comparable corpora and parallel corpora, as in Cappelle and Loock's (2013) study of existential *there is/il y a* 'there is' constructions in English-French translation. Relying on a bilingual comparable corpus, Cappelle and Loock first show that existential *there* is used more freely than its French equivalent *il y a*. The analysis of monolingual comparable corpora of translated and non-translated English and French reveals that existential *there* is under-represented in English translated from French while *il y a* is over-represented in French translated from English. These trends are taken as a sign of source-language interference in translation. The analysis is complemented with parallel corpus data in the two translation directions in order to study inter-language frequency shifts. This approach, which truly integrates insights from contrastive studies, is still uncommon in CBTS but is likely to grow in the future.

As can be seen, the trend towards increased combining of various types of parallel and comparable corpora in contrastive linguistics and translation studies, together with the shared awareness of the fundamental roles played by cross-linguistic contrasts, source-language/source-text influence and other translation-related cognitive, social and technological phenomena in translated texts, are clear signs that cross-fertilization of CBCL and CBTS is now well under way, following several calls for rapprochement in the last twenty years (e.g. Granger, 2003; Olohan, 2004; Xiao, 2010).

4. Overview of the special issue

The present special issue contains five articles. It starts with a review of corpus-based research in contrastive linguistics, followed by four empirical case studies. The five articles were all originally presented at the fifth *Using Corpora in Contrastive and Translation Studies* conference, organized by the Centre for English Corpus Linguistics (University of Louvain) in September 2018.

The focus of the special issue is on the combined use of comparable and parallel corpus data in contrastive studies, a methodology which De Baets *et al.* (2020: 87) refer to as the “Johansson procedure” in recognition of Stig Johansson’s pioneering work. The four corpus-based studies included in the special issue illustrate the two ways in which comparable and parallel corpora can be combined and generally show how the two types of corpus data feed into each other. In their articles, Nádvorníková, Dupont, and Oksefjell Ebeling and Ebeling use multilingual comparable corpora to identify cross-linguistic contrasts, which are then examined in parallel corpora. Axelsson, by contrast, starts with the investigation of translation shifts in a bidirectional parallel corpus, before using a comparable corpus to determine whether these shifts reflect cross-linguistic contrasts. The two approaches make it possible to empirically identify contrasts across languages and to distinguish between translation patterns that mirror cross-linguistic differences between the languages compared, on the one hand, and translation-related phenomena, such as explicitation/implicitation (Nádvorníková) or normalization (Dupont), on the other. They show convincingly that cross-linguistic differences, identified on the basis of solid comparable corpus data, are reflected in parallel corpora. However, they also acknowledge that some of the trends emerging from the parallel corpus data are indicative of more general features of translated language. In doing so, these studies recognize the complex nature of translation and go some way towards reconciling the typical foci of contrastive linguistics (contrasts) and translation studies (features of translated language) (cf. Granger, 2003).

While Nádvorníková and Axelsson examine a single register (fiction), the other two studies address inter- and intra-register variation. In Dupont’s article, two written registers are compared (research articles and newspaper editorials). This approach is in line with recent CBCL research that shows that some cross-linguistic contrasts vary across registers (Neumann, 2013, 2014; Biber, 2014; Lefer and Vogeleer, 2014). This is yet another point of convergence with CBTS, where register variation also features high on the agenda (cf. Kruger and Van Rooy, 2012). Oksefjell Ebeling and Ebeling, for their part, examine an under-researched aspect of register variation in contrastive studies: intra-register variation. Following Egbert and Mahlberg’s (2020) comparative approach to narration and dialogue (speech) in fiction, they set out to determine whether the cross-linguistic contrasts identified on the basis of fiction corpora vary across narrative passages and dialogue.

The special issue starts with an overview of the field of corpus-based contrastive studies by Hilde Hasselgård, from its beginnings in the early 1990s to the present day. She highlights the pioneering role played by Stig Johansson in the major turnaround in contrastive studies brought about by the use of computer corpora. Not only did he specify the contours of the field but he also laid the foundation for bi- and multi-lingual corpus design and methodology. Hasselgård devotes special attention to the notion of *tertium comparationis* and the difficulties it presents, whether it relies on translation equivalence (parallel corpora) or text comparability and assumed cross-linguistic equivalence of linguistic categories (comparable corpora). She traces the development of corpus-based contrastive studies by carrying out a survey of articles published in *Languages in Contrast* from 1998 to 2018. The survey shows a steady increase in corpus-based studies, with a particular focus on discourse, lexis and syntax. It also provides some evidence of language diversification, with a growing number of studies involving languages other than English. Register differentiation and the investigation of phraseology are presented as two emerging trends which will help further widen the scope of the field. Hasselgård also addresses the key challenges for corpus-based contrastive studies identified by Johansson (2012) and distinguishes between those that have been solved and those that have yet to be resolved. The latter include the annotation of bi- and multi-lingual corpora for features other than POS-tagging and lemmatization. Although extremely labour-intensive, semantic and discourse annotation hold great potential for contrastive studies. In the last section of the article, Hasselgård offers a glimpse of promising future directions for the field. These include the compilation and analysis of multimodal

corpora, a stronger focus on contrastive corpus pragmatics and the need to replicate some of the older corpus collections to better reflect current language use.

The first empirical study by Olga Nádvorníková compares the lexical variation of reporting verbs in English, French and Czech. Previous research has shown that the three languages differ markedly in their use of reporting verbs: English and to some extent French have a preference for the neutral reporting verbs – respectively *say* and *dire* – while Czech is characterized by rich variation of semantically more specific verbs. The aim of the study is to provide further empirical evidence of this contrast by carrying out two corpus studies: a contrastive study aimed at analysing the types of reporting verb and their lexical variation, and a translation study aimed at investigating the impact of the differences uncovered on the degree of explicitation/implication in the three languages. The study is based on fiction data extracted from the InterCorp corpus totalling 30 million words (cf. Čermák and Rosen, 2012). Based on a typology that categorizes reporting verbs into five degrees of semantic specificity, the comparable analysis corroborates the results of previous research. It shows that the type/token ratio and hapax/token ratio is significantly higher in Czech than in English and French owing to Czech's tendency to use verbs that convey specific information such as the speaker's emotions or voice quality. The aim of the translation part of the study is to check whether this tendency will result in greater explicitation in translations from English and French to Czech. The results largely confirm this hypothesis but show that the shifts mainly concern the neutral verbs *say* and *dire*. The analysis of all six translation directions shows that explicitation is language-pair specific, which offers further support to related research that has cast doubt on the universality of explicitation. The results also confirm Klaudy and Károli's (2005) asymmetry hypothesis, as explicitation in one direction is only partly counterbalanced by implication in the other direction. The avenues for future research suggested by the author include the need to complement the current study with analyses that not only focus on the verbs but also take into account the rest of the reporting clause, as specification of reported speech is often conveyed by other means, particularly by adjuncts.

The next empirical study, by Karin Axelsson, deals with questions ending in a question mark (QMquestions) in English and Swedish fiction. Research on QMquestions, which include a wide range of structures, has been relatively patchy in contrastive research to date, especially as regards the English-Swedish pair. Axelsson's contribution aims to fill that gap. The study, which relies on the fiction component of the English-Swedish Parallel Corpus, is made up of two complementary parts: a bidirectional parallel corpus analysis and two comparable corpus analyses. In the parallel analysis, Axelsson offers a rich overview of the mutual correspondence (MC) of QMquestion types in English and Swedish, with particular emphasis on translation bias, i.e. cases where non-congruent translations are significantly more frequent in one translation direction than in the other. Three main categories of QMquestion emerge from the analysis: (i) question types with high MC scores (e.g. alternative questions); (ii) question types with relatively high MC scores but a marked translation bias (e.g. wh-interrogatives and non-wh-fragments); (iii) question types with lower MC scores (e.g. declarative questions, wh-fragments and, to a lesser extent, tag questions). The bilingual comparable analysis of original English and Swedish fiction shows that lower MC scores correspond closely to cross-linguistic frequency and structural differences. For instance, wh-fragments and tag questions are more common in English originals than in Swedish originals, while the reverse holds true for declarative questions. The data also reveal that some English wh-fragment structures are not attested in Swedish (e.g. *what about*). Declarative questions differ cross-linguistically as well, with English favouring the use of subject *you* while Swedish typically relies on a range of modal particles in this question type. Axelsson also includes monolingual comparable analyses of original and translated language in her study. These show that the cross-linguistic contrasts uncovered on the basis of parallel and comparable data are reflected in under- and overuse patterns in translated English and translated Swedish, as compared with originals in the same language. For example, wh-fragments and tag questions are underused in translated English, while tag questions are overused in translated Swedish. Purported translation-inherent trends are also identified in translation, such as levelling out and explicitation.

The third empirical study in the special issue, by Maïté Dupont, examines the placement patterns of English and French conjunctive adjuncts of contrast (e.g. *by contrast*,

however, on the other hand) and their role in French-to-English translation, with a view to evaluating the degree of naturalness of translators' syntactic choices. Building on cross-register contrastive studies grounded in Systemic Functional Linguistics (SFL) such as Neumann (2014), Dupont first conducts a bilingual comparable analysis on the basis of corpora of newspaper editorials and research articles in the social sciences, totalling four million words each. The comparable analysis of placement patterns relies on an innovative, SFL-informed taxonomy of five thematic and rhematic positions of conjunctive adjuncts. The results of the Classification and Regression Trees method used in the study show that, although variation across registers plays a significant role in adjunct placement in English and French, language is a better predictor of placement than register, since cross-linguistic differences are stable across editorials and research articles. While conjunctive adjuncts of contrast are typically placed initially or between the topical Theme and the main verb phrase of the clause complex in English, French favours the use of adjuncts of contrast within the main verb phrase of the clause complex. The bilingual comparable approach is complemented with two additional corpus analyses focusing on news. First, Dupont takes a monolingual comparable approach by comparing English news translated from French with original English news. Two differences emerge: (i) a larger proportion of initial uses in translated news, which can be interpreted as normalization, i.e. a manifestation of translators' tendency to opt for the most prototypical position of the target language; (ii) a larger proportion of conjunctive adjuncts used inside the main verb phrase in translated news, which suggests interference from French. Second, positional shifts are analyzed in a French-to-English parallel corpus of news articles. The data reveal that most of the positional shifts mirror the cross-linguistic differences uncovered in the bilingual comparable analysis, indicating that news translators, when applying positional shifts, conform to the syntactic preferences of the target language (here, English).

In the last contribution to this special issue, Signe Oksefjell Ebeling and Jarle Ebeling combine comparable and parallel corpus data to examine intra-register variation in fiction. In particular, they assess the extent to which English-Norwegian contrasts previously identified in the fictional portion of the ENPC hold true when a distinction is made between dialogue and narrative (cf. Egbert and Mahlberg, 2020). Using corpus-driven methods (word and 3-gram lists), Oksefjell Ebeling and Ebeling show that it is useful to distinguish between dialogue and narrative when studying fiction contrastively, as the two sub-registers display different lexico-grammatical features in the two languages investigated. On the basis of frequency word lists, they select two English items that are proportionally more frequent in dialogue than in narrative: *there* and *see*. The reason for choosing these items, in addition to the frequency differences just mentioned, is that *there* and *see* have both been investigated in previous research based on the ENPC, where fiction was studied as a whole (Ebeling, 2000; Øhman, 2006). The analyses across the two sub-registers reveal that there are subtle differences between the use of the two items in dialogue and narrative. For example, the comparable corpus data reveal that the preferred use of *see* in dialogue is in the cognition sense ('understand'), while narrative favours the perception sense ('perceive with eyes'), which in turn affects the translation paradigm of *see* in dialogue and narrative parallel texts. Another noticeable difference between dialogue and narrative, found in the parallel corpus data, is the choice of verb in the Norwegian translations of English existential *there*-clauses. Dialogue is restricted to verbs of existence, while narrative is not, a trend which had gone unnoticed in Ebeling (2000). The study has important implications for future contrastive research based on fiction and possibly on other registers such as news and research articles, as it shows that intra-register variation needs to be taken into account in both comparable and parallel corpus analyses. In addition, it illustrates the added value of replication studies in corpus-based contrastive linguistics, a rare endeavour to date, as well as the challenges replication generates in terms of data extraction and analysis.

Acknowledgements

We would like to thank the reviewers for their generous help and insightful comments on the first versions of the articles included in the special issue. Special thanks also go to Hilde Hasselgård and Signe Oksefjell

Ebeling, former General Editors of *Languages in Contrast*, for trusting us with the guest edition of this issue of the journal and to Laura Penha-Marion for her help in formatting the articles.

References

- Aijmer, K. and Lewis, D. 2017. Introduction. In *Contrastive Analysis of Discourse-Pragmatic Aspects of Linguistic Genres*, K. Aijmer and D. Lewis (eds), 1–9. Cham: Springer.
- Altenberg, B. 1999. Adverbial Connectors in English and Swedish: Semantic and Lexical Correspondences. In *Out of Corpora. Studies in Honour of Stig Johansson*, H. Hasselgård and S. Oksefjell (eds), 249–268. Amsterdam: Rodopi.
- Altenberg, B. and Aijmer, K. 2000. The English-Swedish Parallel Corpus: A Resource for Contrastive Research and Translation Studies. In *Corpus Linguistics and Linguistic Theory*, C. Mair and M. Hundt (eds), 15–33. Amsterdam: Rodopi.
- Baker, M. 1993. Corpus Linguistics and Translation Studies. Implications and Applications. In *Text and Technology. In Honour of John Sinclair*, M. Baker, G. Francis and E. Tognini-Bonelli (eds), 233–250. Amsterdam: John Benjamins.
- Baker, M. 1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target* 7(2): 223–243.
- Becher, V. 2011. When and Why do Translators Add Connectives? A Corpus-based Study. *Target* 23(1): 26–47.
- Bernardini, S. 2011. Monolingual Comparable Corpora and Parallel Corpora in the Search for Features of Translated Language. *SYNAPS* 26: 2–13.
- Biber, D. 2014. Using Multi-dimensional Analysis to Explore Cross-Linguistic Universals of Register Variation. *Languages in Contrast* 14(1): 7–34.
- Bossuyt, T. and Leuschner, T. 2020. WH-ever in German, Dutch and English: A Contrastive Study Showcasing the ConverGENTecorpus. In *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, R. Enghels, B. Defrancq and M. Jansegers (eds), 183–219. Berlin: De Gruyter.
- Cappelle, B. and Loock, R. 2013. Is there Interference of Usage Constraints? A Frequency Study of Existential *there is* and its French Equivalent *il y a* in Translated vs. Non-translated Texts. *Target* 25(2): 252–275.
- Cartoni, B. and Meyer, T. 2012. Extracting Directional and Comparable Corpora from a Multilingual Corpus for Translation Studies. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC '12)*. Istanbul, Turkey. 21–27 May 2012. European Language Resources Association (ELRA). 2132–2137.
- Čermák, F. and Rosen, A. 2012. The Case of InterCorp, a Multilingual Parallel Corpus. *International Journal of Corpus Linguistics* 13(3): 411–427.
- Dai, G. and Xiao, R. 2011. “SL shining through” in Translational Language: A Corpus-Based Study of Chinese Translation of English Passives. *Translation Quarterly* 62: 85–108.
- De Baets, P., Vandevoorde, L. and De Sutter, G. (2020). On the Usefulness of Comparable and Parallel Corpora for Contrastive Linguistics. Testing the Semantic Stability Hypothesis. In *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, R. Enghels, B. Defrancq and M. Jansegers (eds), 85–126. Berlin: De Gruyter.
- Defrancq, B. and Collard, C. 2020. Using Data from Simultaneous Interpreting in Contrastive Linguistics. In *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, R. Enghels, B. Defrancq and M. Jansegers (eds), 159–181. Berlin: De Gruyter.
- Delaere, I. and De Sutter, G. 2017. Variability of English Loanword Use in Belgian Dutch Translations: Measuring the Effect of Source Language, Register, and Editorial Intervention. In *Empirical Translation Studies: New Methodological and Theoretical Traditions*, G. De Sutter, M.-A. Lefer and I. Delaere, 81–112. Berlin: De Gruyter Mouton.
- Denturck, K. 2012. Explicitation vs. Implication: A Bidirectional Corpus-based Analysis of Causal Connectives in French and Dutch translations. *Across Languages and Cultures* 13(2): 211–227.
- Ebeling, J. 2000. *Presentative Constructions in English and Norwegian. A Corpus-Based Contrastive Study*. Oslo: Acta Humaniora.
- Ebeling, J. and Oksefjell Ebeling, S. 2015. An English-Norwegian Contrastive Analysis of Downtoners, more or less. *Nordic Journal of English Studies* 14(1): 62–89.
- Egan, T. 2019. Non-representativeness in Corpora: Perils, Pitfalls and Challenges. *CogniTextes* [Online] 19. Available at <http://journals.openedition.org/cognitextes/1772> [last accessed 24 June 2020].
- Egbert, J. and Mahlberg, M. 2020. Fiction – one Register or two? Speech and Narration in Novels. *Register Studies* 2(1): 72–101.
- Enghels, R., Defrancq, B. and Jansegers, M. 2020. Reflections on the Use of Data and Methods in Contrastive Linguistics. In *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, R. Enghels, B. Defrancq and M. Jansegers (eds), 1–20. Berlin: De Gruyter.

- Frankenberg-Garcia, A. 2019. A Corpus Study of Splitting and Joining Sentences in Translation. *Corpora* 14(1): 1–30.
- Frankenberg-Garcia, A. and Santos, D. 2003. Introducing Compara, the Portuguese-English Parallel Corpus. In *Corpora in Translator Education*, F. Zanettin, S. Bernardini and D. Stewart (eds), 71–88. Manchester: St Jerome.
- Gellerstam, M. 1996. Translations as a Source for Cross-Linguistic Studies. In *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*, K. Aijmer, B. Altenberg and M. Johansson (eds), 53–62. Lund: Lund University Press.
- Granger, S. 2003. The Corpus Approach: A Common Way forward for Contrastive Linguistics and Translation Studies. In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, S. Granger, J. Lerot and S. Petch-Tyson (eds), 17–29. Amsterdam: Rodopi.
- Hansen-Schirra, S., Neumann, S. and Steiner, E. 2012. *Cross-Linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. Berlin: De Gruyter.
- Hasselgård, H. 2014. Discourse-Structuring Functions of Initial Adverbials in English and Norwegian News and Fiction. *Languages in Contrast* 14(1): 73–92.
- Ibrahim, R.K. 2015. Translation Oriented Corpus-Based Contrastive Linguistics. *Babel* 61(3): 381–393.
- James, K. 1980. *Contrastive Analysis*. London: Longman.
- Johansson, S. 1998. On the Role of Corpora in Cross-Linguistic Research. In *Corpora and Cross-Linguistic Research. Theory, Methods, and Case Studies*, S. Johansson and S. Oksefjell (eds), 3–24. Amsterdam: Rodopi.
- Johansson, S. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.
- Johansson, S. 2011. A Multilingual Outlook of Corpora Studies. In *Perspectives on Corpus Linguistics*, V. Viana, S. Zyngier and G. Barnbrook (eds), 115–129. Amsterdam: John Benjamins.
- Johansson, S. 2012. Cross-Linguistic Perspectives. In *English Corpus Linguistics: Crossing Paths*, M. Kytö (ed), 45–68. Amsterdam: Rodopi.
- Kenning, M.-M. 2010. What are Parallel and Comparable Corpora and how can we Use them? In *The Routledge Handbook of Corpus Linguistics*, A. O’Keeffe and M. McCarthy (eds), 487–500. London: Routledge.
- Kenny, D. and Sathachai, M. 2018. Explicitation, Unique Items and the Translation of English Passives in Thai Legal Texts. *Meta* 63(3): 604–626.
- Klaudy, K. and Károly, K. 2005. Implication in Translation: Empirical Evidence for Operational Asymmetry in Translation. *Across Languages and Cultures* 6(1): 13–28.
- Kotze, H. 2019. Converging *what* and *how* to Find out *why*: An Outlook on Empirical Translation Studies. In *New Empirical Perspectives on Translation and Interpreting*, L. Vandevoorde, J. Daems and B. Defrancq (eds), 333–371. New York: Routledge.
- Kruger, H. and Van Rooy, B. 2012. Register and the Features of Translated Language. *Across Languages and Cultures* 13(1): 33–65.
- Lauridsen, K.M. 1996. Text Corpora and Contrastive Linguistics: Which Type of Corpus for which Type of Analysis? In *Languages in Contrast. Papers from a Symposium on Text-based Cross-Linguistic Studies*, K. Aijmer and B. Altenberg (eds), 63–71. Lund: Lund University Press.
- Lefer, M.-A. and De Clerck, M. forthcoming. L’apport des Corpus Intermodaux en Lexicologie Contrastive: Étude Comparative de la Traduction Écrite et de l’Interprétation Simultanée des Séquences de Noms. In *Autour de l’Énonciation, de la Lexicologie, de la Morphophonologie et de la Contrastivité: Langues, Discours, Textes et Corpus*, S. Hanote and R. Nita (eds). Rennes: Presses Universitaires de Rennes.
- Lefer, M.-A. and Vogeeler, S. (eds). 2014. *Genre- and Register-Related Discourse Features in Contrast*. Special issue of *Languages in Contrast* 14(1).
- 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.
- Macken, L., De Clercq, O. and Paulussen, H. 2011. Dutch Parallel Corpus: A Balanced Copyright-Cleared Parallel Corpus. *Meta* 56(2): 374–390. <https://doi.org/10.7202/1006182ar>
- Magnifico, C. and Defrancq, B. 2020. Norms and Gender in Simultaneous Interpreting: A Study of Connective Markers. *Translation and Interpreting* 12(1): 1–17.
- Marco, J. 2019. The Translation of Food-Related Culture-Specific Items in the Valencian Corpus of Translated Literature (COVALT) Corpus: A Study of Techniques and Factors. *Perspectives* 27(1): 20–41.
- Mastropierro, L. 2018. Key Clusters as Indicators of Translator Style. *Target* 30(2): 240–259.
- Mauranen, A. 1999. Will ‘Translationese’ Ruin a Contrastive Study? *Languages in Contrast* 2(2): 161–186.
- McEnery, T., Xiao, R. and Yukio, T. 2006. *Corpus-Based Language Studies. An Advanced Resource Book*. London: Routledge.
- Neumann, S. 2013. *Contrastive Register Variation. A Quantitative Approach to the Comparison of English and German*. Berlin: De Gruyter Mouton.
- Neumann, S. 2014. Cross-Linguistic Register Studies: Theoretical and Methodological Considerations. *Languages in Contrast* 14(1): 35–57.

- Øhman, B.I. 2006. *An SFG Perspective on the Polysemy of See: A Corpus-Based Contrastive Study*. Master's Dissertation, University of Oslo.
- Olohan, M. 2004. *Introducing Corpora in Translation Studies*. London: Routledge.
- Pavesi, M. 2013. *This and that in the Language of Film Dubbing: A Corpus-Based Analysis*. *Meta* 58(1): 103–133.
- Pérez Blanco, M. 2016. The Construction of Attitudinal Stance: A Corpus-Based Contrastive Study of Negative Evaluative Adjectives in English and Spanish Opinion Discourse. *Languages in Contrast* 16(1): 31–53.
- Ramón García, N. 2002. Contrastive Linguistics and Translation Studies Interconnected: The Corpus-based Approach. *Themes in Translation Studies* 1. Available at <https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/27> [last accessed 24 June 2020].
- Rozumko, A. 2016. Adverbs of Certainty in a Cross-Linguistic and Cross-Cultural Perspective – English-Polish. *Languages in Contrast* 16(2): 239–263.
- Salkie, R. 2007[1997]. Naturalness and Contrastive Linguistics. In *Corpus Linguistics*, W. Teubert and R. Krishnamurthy (eds), 336–351. London: Routledge.
- Teich, E. 2003. *Cross-Linguistic Variation in System and Text*. Berlin: Mouton De Gruyter.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: Benjamins.
- Winters, M. 2013. German Modal Particles – from Lice in the Fur of our Language to Manifestations of Translators' Styles. *Perspectives* 21(3): 427–445.
- Xiao, R. 2011. Word Clusters and Reformulation Markers in Chinese and English. Implications for Translation Universal Hypotheses. *Languages in Contrast* 11(2): 145–171.
- Xiao, R. (ed). 2010. *Using Corpora in Contrastive and Translation Studies*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Xiao, R., McEnery, T. and Qian, Y. 2006. Passive Constructions in English and Chinese: A Corpus-Based Contrastive Study. *Languages in Contrast* 6(1): 109–149.

Address for correspondence

Sylviane Granger
 Université catholique de Louvain
 Centre for English Corpus Linguistics
 Collège Erasme
 Place Blaise Pascal 1
 B-1348 Louvain-la-Neuve
 Belgium
 sylviane.granger@uclouvain.be
<https://orcid.org/0000-0002-0047-2142>

Co-author information

Marie-Aude Lefer
 Université catholique de Louvain
 Centre for English Corpus Linguistics
 marie-aude.lefer@uclouvain.be