Corpora et comparatio linguarum: **Textual and contextual perspectives**

Signe Oksefjell Ebeling, Hilde Hasselgård

University of Oslo (Norway)

1.   Introduction

This collection of papers arose out of the contrastive pre-conference workshop at the 38th ICAME[1] conference organized by Charles University in Prague in May 2017. The first part of the title of this issue of BeLLS, and indeed the title of the workshop, was inspired by the overall conference title: *Corpus et Orbis: Interpreting the World through Corpora.* The workshop theme – textual and contextual perspectives – was chosen in recognition of the fact that the conference took place in the home university of the Prague School and Functional Sentence Perspective. We were fortunate to be able to include a keynote presentation in the workshop by a prominent representative of this school, Prof. Libuše Dušková. Her own work reflects two of the main concerns of Functional Sentence Perspective: the close association between syntactic form and information structure and the text-based comparison of languages (Dušková 2015). Corpus-based contrastive linguistics is both related to and inspired by the text-based language comparison that existed before multilingual corpora, as acknowledged by Johansson (2009).

The call for papers invited scholars to make use of corpora to report on textual and contextual matters in a cross-linguistic perspective. As pointed out by Johansson (2011), one of the great advantages of multilingual corpora in contrastive linguistics is that they can make "possible a comparison of language use in context. We can compare not just structures, but their conditions of use" (2011: 125). The papers in this collection demonstrate that the contextual perspective may be taken at many levels of linguistic analysis, from the interpretation of single lexical items to the study of information structure. The contextual perspective is evident as well as inevitable in a cross-linguistic study of the FSP of English and Czech (Dušková), cohesive chains (Kunz & Lapshinova-Koltunski), and the use of marked theme (Rørvik & Monsen). Moreover, particular syntactic and phraseological constructions can only be defined, and thereby studied, by reference to their contexts (see studies by Ebeling, Hasselgård, and Ström Herold & Levin). The interpretations of individual lexical items, such as the Swedish verbs *skall* (Aijmer) and *orka* (Johansson & Nordrum), the English *fail to* (Egan), and the postfix *-pak* (Šebestová & Malá) are closely connected with the contexts in which they occur – and moreover, the translators are likely to draw on the

---

[1] International Computer Archive of Modern and Medieval English (`http://clu.uni.no/icame/`).

context in order to find appropriate counterparts in cases where the target language lacks a direct equivalent.

2. Contents of the volume

All the papers in this collection compare English with at least one other language on the basis of parallel (translation) or comparable corpora. The languages studied, in addition to English, are Czech, German, Norwegian and Swedish.

Libuše Dušková's paper offers a critical account of the use of parallel corpora consisting of original and translated texts. The main pitfalls are associated with cross-linguistic differences related to information structure in English and Czech. The three aspects investigated − linear ordering of clause elements, FSP structure and distribution of communicative dynamism − are indeed shown to pose challenges in determining adequate translation counterparts. These challenges notwithstanding, Dušková maintains that the use of parallel texts "is irreplaceable insofar as it is the only methodology that provides expression of the same content worded in different languages" (p. 5).

Kerstin Kunz and **Ekaterina Lapshinova-Koltunski** look at lexical cohesion and chains of coreference in a study of four spoken and written registers of German and English: fiction, essays, interviews, and popular science. Chains of coreference are taken to reveal the development of discourse topics. A contribution of the study is that it considers the two types of cohesion together, in looking at the intersections of lexical cohesion and coreferential chains. The study uncovers cross-linguistic contrasts: In particular, there are more overlapping antecedents and more intersections in German than in English but the number of overlapping anaphors is higher in English than in German. In addition there are cross-register differences within each language. For example, fictional texts are distinct from other registers in both languages, though there are cross-linguistic differences. On the other hand, popular science texts show cross-linguistic similarities as regards chain intersection.

Sylvi Rørvik and **Marte Monsen** investigate the use of marked themes in English and Norwegian within the field of didactics. Using material from research articles written in L1 English and Norwegian and L2 English (by L1 Norwegian speakers), the authors aim to uncover potential contrastive differences (L1 vs. L1) in order to inform novice (L2) writers of good practices with regard to textual features. Rørvik and Monsen demonstrate that, although there are some significant contrastive differences both in terms of realizations and meanings of marked themes in didactics articles, the L2 writers are generally shown to adapt to English, discipline-specific discourse conventions in these respects.

Drawing on material from the English-Norwegian Parallel Corpus+, **Signe Oksefjell Ebeling** explores the cross-linguistic congruence of two stance frames: *it BE ADJ that* and *det VÆRE ADJ at*. The results indicate that, although there is a lot of similarity between the two languages in the use of these frames (55% intertranslatability, or congruence in translation), 45% non-congruence is noted, i.e. overt but formally different correspondences. The degree of congruence is found to depend on the type of attitude/evaluation expressed by the frames. Moreover, both languages are shown to have a number of other expressions of attitudinal stance at their disposal.

Hilde Hasselgård studies sentence-initial indefinite subjects in English and Norwegian. Since indefiniteness is associated with new information, such subjects appear to violate the information principle that co-exists with the SVO principle in both languages. The English-Norwegian Parallel Corpus is used for comparing original texts as well as originals and their translations. Both parts of the study indicate that English is more tolerant of indefinite subjects than Norwegians. However, certain contexts are favourable to indefinite

subjects in both languages, particularly so-called 'bare presentatives' which include a verb of existence/appearance and a place adverbial, generic sentences, and sentences with an indefinite NP in object position. The fact that more changes are made to indefinite subjects in translation from English into Norwegian than vice versa is related to the stronger influence of the light-subject constraint in Norwegian.

Jenny Ström Herold and **Magnus Levin** use a new resource – the Linnaeus University English-German-Swedish parallel corpus (LEGS) – to explore German and Swedish correspondences (translations and sources) of English supplementive *ing*-clauses, a clause type that lacks a productive equivalent in the target languages. It is shown that coordination is by far the most frequent correspondence in both languages, reflecting the compact and semantically indeterminate nature of supplementive *ing*-clauses. Other major correspondence types include subordination, main clause and prepositional phrase. Main clauses are found to be more frequent correspondences in German than in Swedish, which is attributed to the fact that there seems to be an increasing preference for parataxis (rather than hypotaxis) in German overall. Ström Herold and Levin also note some instances of explicitation in both the German and Swedish translations.

Karin Aijmer shows how translations can shed light on multifunctional expressions, and how their interpretation must be informed by the context. The Swedish modal auxiliary *ska/ll* is studied through its English translations found in the English-Swedish Parallel Corpus. Since *ska/ll* is known to differ markedly in meaning and use from its English cognate, it is no surprise that *shall* is an infrequent translation correspondence. However, the analysis reveals a wide array of other correspondences, thus displaying the multifunctionality of *ska/ll*. Apart from its most frequent use in future constructions (where it typically corresponds to *will*), *ska/ll* is often found in performative uses associated with authority and obligation. *Ska/ll* is also involved in other types of speech acts, such as offer, suggestion and advice, where the translations indicate that the imposition on the hearer is weakened.

Mats Johansson and **Lene Nordrum** investigate the Swedish auxiliary *orka* through the lens of its English correspondences in parallel corpora. Since English does not have a straightforward equivalent of *orka*, the various correspondences reveal its meaning components: most importantly, participant-internal ability and sufficient physical or mental strength/energy. The sufficiency component distinguishes it from other ability verbs such as *kunna* ('be able to'). The analysis also finds that *orka* tends to occur in non-assertive, often negative, contexts, which is related to the fact that it is more important to specify sufficiency in the case of unrealized events. Drawing on van der Auwera and Plungian's (1998) *Modality's semantic map*, the authors argue that sufficiency should be regarded as a layer of modality.

Thomas Egan investigates the English construction FAIL TO and its Norwegian correspondences in the English-Norwegian Parallel Corpus. Since Norwegian does not have a direct counterpart to FAIL TO, the correspondences are expected to reveal grammaticalized uses of the expression, particularly in contexts where FAIL TO does not imply effort, duty or expectations on the part of the Subject. FAIL TO is found to be much more frequent in English source texts than in translations from Norwegian, which is related to the absence of such a construction in Norwegian and the relative unlikelihood of translating a default negator such as *ikke* ('not') by a more wordy construction, such as FAIL TO. However, *ikke* is a relatively common translation of FAIL TO, both on its own and in combination with a verb meaning roughly 'manage'. Almost half of the instances of FAIL TO in translations have a negative element such as *ikke* or *ingenting* ('nothing') as their sources. The study thus shows that FAIL TO, at least in some of its uses, can be described as a periphrastic negative.

Using the Czech-English part of the InterCorp, **Denisa Šebestová** and **Markéta Malá** explore the discourse functions of the Czech postfix *-pak*. A number of both primary and

secondary functions are uncovered through an analysis of its English translation counterparts, lending support to the authors' assumption that such counterparts can indeed be used to shed light on this multifunctional postfix. Expressions ending in *-pak* are found to have the primary functions of marking epistemic modality, voicing an appeal, and marking a change in the speaker's previous assumption. In addition, *-pak* expressions are shown to have the ability to mark other pragmatic meanings, including politeness and textual cohesion.

3.   Acknowledgements

References

Dušková, L. 2015. *From Syntax to Text: the Janus Face of Functional Sentence Perspective*. Prague: Karolinum Press.

Johansson, S. 2009. Multilingual corpora: Possibilities and limitations. Unpublished manuscript for a lecture given at Charles University, Prague.

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.

van der Auwera, J. and Plungian, V. A. 1998. Modality's semantic map. *Linguistic Typology* 1998(2): 79–124.