

Cross-linguistic studies at the interface between lexis and grammar

Karin Aijmer, University of Gothenburg and Hilde Hasselgård, University of Oslo

This special issue of the *Nordic Journal of English Studies* comprises a selection of papers that were first presented at a pre-conference workshop on corpus-based contrastive studies at the 34th ICAME conference in Santiago de Compostela in May 2013. They all concern English in comparison with at least one other language, collectively covering languages of Germanic, Romance, Slavic and Baltic origin. The topic of the workshop was “Cross-linguistic studies at the interface between lexis and grammar”. This prompted explorations of linguistic phenomena ranging from single words that are exponents of a grammatical category to collocation and patterns of discourse.

Corpus-based contrastive studies tend to use functional models of linguistics for interpretation; the present collection is no exception. A common feature of functional linguistics is the reluctance to make a sharp distinction between lexis and grammar, and but instead, in the words of Halliday (1994: 15), to see the two as “different ends of the same continuum”. Such a continuum is often highlighted in cross-linguistic comparisons based on translation corpora, as they typically bring out *paradigms of correspondence* (Johansson 2007: 23; Hasselgård 2012: 1549), with members of the paradigms belonging to different form classes. An example is Johansson’s study of *seem* and its Norwegian correspondences (Johansson 2007: 117 ff), which shows that sources and translations of *seem* include modal auxiliaries, modal particles and disjunct adverbials, in addition to congruent verb correspondences; see examples (1)-(4), from the English-Norwegian Parallel Corpus.

- (1) “I don’t *seem* to remember your ever giving Molly much of a chance,” said Frederick. (DL1)
“Jeg *kan* ikke huske at du noensinne har gitt Molly noen særlig sjanse,” sa Frederick. (DL1T)
Lit: ‘I can not remember that you ever...’

Aijmer, Karin and Hilde Hasselgård. 2015. “Cross-linguistic studies at the interface between lexis and grammar.” *Nordic Journal of English Studies* 14(1):1-8.

2 Karin Aijmer and Hilde Hasselgård

- (2) Parents *seem* to have no control over children these days. (PDJ3)
Foreldre har *visst* liten kontroll over barna sine nå for tiden.
(PDJ3T)
Lit: 'Parents have PARTICLE little control over children theirs
nowadays'
- (3) It was, and his whole body *seemed* to calm down ... (FC1T)
Det gjorde den, og hele kroppen hans falt *liksom* til ro, ... (FC1)
Lit: '... and whole the.body his fell somehow to peace.'
- (4) My hunger made the world *seem* bluish. (BO1)
Jeg var så sulten at verden *virket* blålig. (BO1T)
Lit: 'I was so hungry that the world seemed bluish'

Examples (5) and (6) show that a single verb may correspond to multi-word units expressions too. More importantly, they highlight the fact that units of meaning are "much more extensive and varied than is seen in a single word" (Sinclair 2004: 39).

- (5) Most people, *it seemed to me*, took their turn to be football.
(DF1)
Jeg hadde inntrykk av at de fleste opplevde å være fotball av og
til. (DF1T)
Lit: 'I had impression of that the most [people] experienced to be
football now and then'
- (6) Mama *seemed* to realize it too, ... (HW1T)
Det var som om mamma så det hun også, ... (HW1)
Lit: 'It was as if Mama saw it herself too'

As these examples demonstrate, parallel corpora are an ideal testbed for exploring the interface between lexis and grammar: they highlight the fact that similarity of form need not entail similarity of meaning and vice versa (see also Ebeling & Ebeling 2013: 213 ff).

The studies contained in the present volume explore the lexis-grammar interface in different ways: those by Egan & Graedler and

Rabadán discuss collocations of verbs and adverbials. They provide ample evidence for Halliday's (2004: 9) assertion that "languages vary with respect to the 'division of grammatical labour' among the ranks. In particular, certain languages do relatively more grammatical work at group (and clause) rank, whereas other languages do relatively more work at word rank."

Egan and Graedler explore the well-known difference between languages in the coding of manner and path in motion verbs. The study uses a subcorpus of the Oslo Multilingual Corpus (OMC) with Norwegian original texts and English and French translations. The relevant collocations are identified from searches for the compound prepositions *ut av* ('out of') and *inn i* ('into') and their correspondences. The cross-linguistic picture is more complex than might be expected from the distinction between path-framed and satellite-framed languages (Talmy 2000): the study finds no categorical difference between the three languages with respect to types of framing insofar as they all make use of the same range of constructions to code boundary-crossing self-motion events. However, the two types of framing differ as to their proportions, with the English translations resembling the Norwegian originals more than the French translations do.

Rabadán's study focuses on the colligations of the temporal adverbs *still* and *already* with a range of tense/aspect combinations with the aim of exploring some of the ways in which the grammars of English and Spanish encode aspectual transitions. The analysis is based on the parallel corpus P-ACTRES, English originals with Spanish translations. The corpus data reveal great differences between the two adverbs. *Still + verb phrase construction* is often translated into *seguir + gerund* and (NEG) *dejar de + infinitive*, as alternatives to '*aún / todavía*' + *verb phrase constructions*. There is often no direct correspondence of *still* in the translation, but the meaning is preserved in the Spanish verb form with its rich tense/aspect morphology, thus making aspectual adverbs superfluous. Spanish translations of constructions with *already*, however, tend to be more similar to the originals, and often contain the adverb *ya*.

Sinclair (2004: 29 f.) hypothesized that "the notion of a linguistic item can be extended, at least for English, so that units of meaning are expected to be largely phrasal". The next three papers in this special issue investigate the validity of this hypothesis beyond English: moving

the focus closer to the lexical end of the lexicogrammatical continuum, they present cross-linguistic comparisons of phraseological patterns.

Ebeling and Ebeling's paper compares two similar-looking binomials in English and Norwegian *more or less* / *mer eller mindre* using an extension of the fiction part of the English Norwegian Parallel Corpus, the ENPC+. In spite the formal similarity of the two expressions, they differ in frequency, and are not always each other's translations due to different repertoires of syntactic functions and thereby meanings. Such differences are made visible in the correspondence patterns of the expressions. The paper introduces the notion of *reverse mutual correspondence* to complement Altenberg's (1999) measure of mutual translatability of two linguistic constructions.

Brems's study concerns another type of binomial, namely size noun constructions of the type *heap(s) of N* and *bunch of N* and their French counterparts in a variety of corpora. The French expressions are identified through dictionary data, and the constructions are compared mainly by means of monolingual corpora of English and French, but also translation data from the OPUS corpus. An important conclusion is that despite structural similarity between an English and a French construction, there may be differences in the frequency of their use and only partial constructional equivalence, thus making translation equivalence, too, only partial.

Bondi and Diani use comparable corpora of blog texts to identify semantic sequences of subjective evaluation in English and Italian, starting from first person pronouns. The sequences are classified as 'basic' (evaluation + entity/process evaluated), 'framed' (evaluation + self-attributive framework), and 'argumentative' (writer expressing (dis-)agreement with somebody). The difference in pronominal systems creates divergences in key expressions (*mi piace* vs. *I like*). However, a major finding is that evaluative sequences in English and Italian blogs express very similar meanings in spite of lexical and syntactic differences. This insight is believed to be helpful in future contrastive studies.

The third section of the issue tackles expressions of modality, an area in which languages display great diversity as regards the lexicogrammatical repertoire and the usage conventions for the various expressions.

Lewis investigates expressions of obligation in a comparable corpus of political speeches in English and French. The focus is on the modal and semi-modal verbs *must*, *need to*, *have to*, *devoir* and *falloir*. The study reveals differences in formality across the expressions, particularly with *must* and *devoir* turning up in more formal contexts. In addition these expressions are markedly more frequent in political speeches than in spoken English and French in general. In spite of different realizations, there are salient similarities in the ways in the expression of obligation is handled in political speeches in both languages.

Aijmer analyses the Swedish modal particle *väl* by studying how its meanings are mirrored in its translations into English and French on the basis of parallel corpora. It is shown that there are no easy ways of translating *väl* into the other language. There were a large number of non-recurrent correspondences in the corpora studied suggesting that it does not have a direct translation. Moreover the translator often omitted *väl* when this was possible. The English and French correspondences highlight semantic aspects or sub-functions of *väl* such as certainty, hedging, asking for a response, appeal, argumentative uses where the speaker pretends to know best. Some of the translations focus on the modal functions of *väl* to express a high or low degree of certainty while others highlight the function to appeal to the hearer as 'the best knower'. The translations also show that *väl* can express the speaker's negative attitudes such as reproach or annoyance.

Usoniené et al. extend earlier contrastive studies of the English adverbials *actually* and *in fact* to Lithuanian. Using a bidirectional English-Lithuanian parallel corpus based on the design of the English Norwegian Corpus the authors investigate how the functions of *actually* and *in fact* are reflected in Lithuanian. As can be seen from their translation paradigms both *in fact* and *actually* have a large number of translation correspondences. It is shown that a large number of the translations are shared, indicating the semantic closeness between the items. The prototypical translations both of *in fact* and *actually* are seen to belong to the semantic domains of TRUTH and REALITY, which dominate over domains such as FACT or ACT. In order to show the semantic and pragmatic potential of the translation correspondences of the markers their functions are also analysed in Lithuanian academic discourse and compared with the functions that they traditionally have.

Finally, three papers discuss diverse aspects of discourse structure, thus moving the focus to the textual metafunction of language. They show similarities and differences between languages as regards clause combination, cohesion and thematic structure.

Malá and Saldová compare the use of participial forms as a condensation device in English and Czech. In English participles are used systematically both as adverbials and modifiers of nouns while in Czech the participial construction is obsolete or non-existent. The fiction texts of the English-Czech section of the multilingual parallel corpus *InterCorpus* are used to investigate the translation correspondences between the languages and to draw attention to some general features attributed to English participial clauses. To begin with the Czech correspondences were sorted into congruent and divergent ones. It is shown that the Czech counterparts highlight the analytic nature of English where the verbal meaning is dissociated into the finite verb containing the grammatical information and the participle which contains the semantic core of the construction. For English postmodifying clauses a subordinate finite clause appears to be the preferred translation alternative (the hypotactic choice). The adverbial participial clauses on the other hand display a preference for finite coordinate clauses (the paratactic choice). The possible non-equivalence between the participial clause and its finite counterpart concerns the fact that the subordinate non-finite clause makes it possible to present a proposition as backgrounded.

The study by Kunz and Lapshinova-Koltunski analyses and interprets contrasts in three types of cohesion—co-reference, viz. co-reference, substitution and conjunction in English and German. The three types are examined in combination in the GECCO corpus, which has been annotated with information on cohesive devices. The main research question is whether contrasts are more pronounced between registers independent of language or whether more differences are identified in one and the same register between English and German. It is shown that substitution plays a minor role in all registers across languages and that high frequencies can be found for co-reference and conjunction. Registers can be identified across languages on the basis of particular cohesive subtypes. Fiction in both English and German originals for example stands out in terms of the heterogeneous distribution of cohesive subtypes. Correspondence analysis can provide information on the

correlation between languages and register. It suggests that both spoken English and spoken German have different preferences from writing. Spoken registers have for example a preference for cohesive device with an interpersonal or rhetorical function. However the differences between languages are generally greater than between registers.

We are beginning to see many different types of bilingual corpora. Lavid and Moratón use a bilingual comparable corpus of Letters to the editor to contrast expressions of stance and engagement in the two languages placed in thematic position. Another aim is to investigate if such expressions are used in the same way in other newspaper genres such as news reports and editorials. Several different thematic choices are investigated. Hyland's model of intersubjective positioning was chosen in order to classify different types of resources that the speakers use to express their positions. However in order to be able to compare English and Spanish themes the authors break up the thematic field into the Inner Thematic Field and the Outer Thematic Field. The analysis revealed a number of language-specific preferences in the use of expressions of stance and engagement.

In recent linguistic theories it has been argued that lexical items should be categorized as patterns or constructions (e.g. Fried & Östman 2004; Sinclair 2004; Ebeling & Ebeling 2013), i.e. to units of meanings or combinations of form and function. All the contributions presented in this special issue have shown the fruitfulness of extending the contrastive analysis in the direction of constructions. Moreover words or constructions do not get their meaning in a vacuum but in the larger contexts. We need to specify for example whether the cross-cultural differences are found in speech or writing (or both) and consider the importance of genre (cf. Lefer & Vogeleer 2014).

The call for papers invited contributors to the workshop to explore the possibility of comparing constructions rather than single words in different languages. As Rosa Rabadán points out in her paper, there are clear benefits in "contrasting (grammatical) constructions rather than single elements, as it helps reveal cross-linguistic associations that are not readily accessible". We whole-heartedly support this view, and believe the present collection of papers demonstrates that this is the way forward for corpus-based contrastive studies.

References

- Altenberg, Bengt. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In H. Hasselgård and S. Oksefjell (eds), *Out of Corpora. Studies in Honour of Stig Johansson*. Amsterdam: Rodopi, 249-268.
- Ebeling, Jarle & Signe Oksefjell Ebeling. 2013. *Patterns in Contrast*. Amsterdam/Philadelphia: Benjamins.
- Fried, Miriam & Jan-Ola Östman (eds). 2004. *Construction Grammar in a Cross-Language Perspective*. Amsterdam: Benjamins.
- Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. 2nd edn. London: Arnold.
- Halliday, M.A.K. 2004. *An Introduction to Functional Grammar*. 3rd edn., revised by Christian M.I.M. Matthiessen: London. Arnold.
- Hasselgård, Hilde. 2012. Cross-linguistic differences in grammar. In Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Wiley-Blackwell, 1547-1554. DOI:10.1002/9781405198431.wbea10290.
- Johansson, Stig. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: Benjamins.
- Lefer, Marie-Aude & Svetlana Vogeleer (eds). 2014. *Genre- and Register-related Discourse Features in Contrast*. Special issue of *Languages in Contrast* (14:1). Benjamins.
- Sinclair, John. 2004. *Trust the Text: Language, Corpus, and Discourse*. London: Routledge.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics*. Cambridge MA: MIT Press.