Introduction – Languages in contrast 20 years on

Lene Nordrum, Lund University, Signe Oksefjell Ebeling and Hilde Hasselgård, University of Oslo

This special issue of the *Nordic Journal of English Studies* comprises papers from the symposium *Languages in Contrast* held in Lund 5 December 2014 in celebration of the 20th anniversary of the Nordic Parallel Corpus Project which led to the English-Norwegian Parallel Corpus (ENPC), the Swedish-English Parallel Corpus (ESPC) and the Finnish-English Contrastive Corpus Studies Project (FECCS). When they were first compiled, the parallel corpora opened up new possibilities for empirical, corpus-based questions about language relationships and translations (see Johansson 2007, 2012: 1). The catalyst for these possibilities was in many ways the structure of the parallel corpora: they are bidirectional translation corpora which facilitate comparisons of comparable original texts in two languages, original and translated text in one language and original text in one language and their translations into another language. The fact that original English texts were to a great extent common to the three original corpora also allowed some degree of comparison of translations (into different languages) of the same text.

The ideas and thinking that gave rise to the parallel corpus project are described by Jarle Ebeling in an introductory overview opening this special issue. In this overview, Ebeling takes stock of the background and the early history of parallel corpora, in particular the English-Norwegian Parallel Corpus and the English-Swedish Parallel Corpus. The title of his paper reflects the fact that the development of such corpora and the techniques for exploring them in linguistic analysis has changed the field of contrastive linguistics – in the words of Stig Johansson (2012) we can now talk about "contrastive linguistics in a new key".

In the papers that follow, the tonal environment of this new key is illustrated, examined and discussed in two main sections: *Exploring contrastive methods* and *Contrasting word meaning and use*. By way of a series of small-scale investigations, the papers in the first section test the methodological power of the principal tone of the key, the parallel corpus, and point to methodological strenghts and potential new uses as

well as future challenges. The section includes papers by Johan van der Auwera and Evie Coussé, Signe Oksefjell Ebeling, Hilde Hasselgård, Thomas Egan and Åke Viberg. In the second section, parallel corpora are used to study the meaning of words in one language by consulting their correspondences in another. The section includes one paper by Karin Aijmer and one by Mats Johansson and Lene Nordrum – the latter paper inspired by Bengt Altenberg's talk at the symposium in Lund. The papers underline Stig Johansson's (2007: 28) insight that multilingual corpora have the advantage of making visible meanings that are difficult to access in monolingual corpora, and thus make a case for parallel corpora in semantic and pragmatic studies beyond those with strictly contrastive concerns.

In what follows, we present the papers in their order of appearance. In the first paper in the section *Exploring contrastive methods*, Johan van der Auwera and Evie Coussé use three different methods to shed light on the similarities and differences between English *such* and Swedish *sådan*, which they describe as 'similatives'. Through this triangulation of combining language-specific scholarship, analysis of comparable and parallel corpora, they demonstrate that while the *possible* uses of *such* and *sådan* are similar, their *actual* uses are not so to the same extent. Both the comparable and parallel corpus analyses are carried out on the basis of the ESPC. English *such* is almost exclusively used adnominally, while Swedish *sådan* is also frequently used (pro)nominally. Identifying and intensifying uses are common for both, but the intensifying use is much more prominent with *such*. Further, the recognitional use is only found with *sådan*. The similarity use is well-established for both *such* and *sådan*, but *such* sees more competition from *like/sort/kind* than *sådan* from *sort/typ*. van der Auwera and Coussé conclude that *such* and *sådan* are neither the same nor different; rather, they are similar.

The paper by van der Auwera and Coussé demonstrates how parallel corpora can serve as a crucial component in studies combining different materials and methodologies. The particular advantage of parallel corpora is its unique ability of highlighting patterns of contrastive pragmatics. A problem that is often raised concerning such patterns, however, is the question of size. Can we in fact trust that small parallel corpora, such as the ENPC and ESPC, give rise to representative patterns? Or, put differently, to what extent does corpus size matter? Signe Oksefjell Ebeling addresses this question in her paper. She uses an

extended version of the fiction part of the ENPC, the ENPC+, to revisit three earlier ENPC case studies and concludes that corpus size alone is perhaps not as decisive as we may have feared, but at the same time, the corpus structure and design in general are far from trivial matters. The first of Ebeling's studies suggests that reliable results can be arrived at on the basis of small-size corpora, since the replica study based on a larger corpus produces virtually the same results as the original study, whereas in the second case study, the time lag between the some of the texts in the ENPC and ENPC+ seems to be a major factor, as a language change appears to be taking place. Other factors contributing to conflicting results between the original and the replica emerge in the third study, including individual style/preference on the part of the author/translator and the length of corpus texts.

Hilde Hasselgård breaks new ground in taking a colligational pattern based on function words as her starting point, the sequence "*the* N1 *of the* N2", (e.g. *the end of the day*). The paper has a double aim: to discover recurrent patterns in the lexical and semantic make-up of this colligational pattern in English and its correspondences in Norwegian, and to explore the possibilities of carrying out a contrastive study on the basis of function words. The English pattern is found to select the N1 more systematically than the N2, and the Norwegian correspondences are more frequently congruent in translations than in sources. Common non-congruent correspondences include compound nouns, *s*-genitives and expressions where the English N1 corresponds to a Norwegian adverb. Individual (N1) lexemes may have their preferred type of correspondence. Hasselgård concludes that the bottom-up approach represented by the function-word starting point can be a fruitful basis of cross-lingusitic study.

In his article entitled "Contrasting translations", Thomas Egan discusses the advantages of using translations from the same source text into several languages as well as the weaknesses of the more traditional contrastive approach of comparing originals and translations to compare expressions in two languages. The method of using translations for contrastive observations has been referred to as 'parallel translations' by Johansson (2007: 31), but has not been extensively explored in the literature so far. To support his argument for the advantages of parallel translations, Egan draws on material from a version of the ENPC and ESPC only including the English original texts that the corpora have in

common, as well as the Oslo Multilingual Corpus (OMC). For example, on the basis of Norwegian source texts with translations into English and French it is shown that English and French use different means to encode temporal and perceptual [throughness]. Based on Altenberg (1999), Egan further proposes a formula for measuring the degree of mutual correspondence between items in translations from the same source into different languages. Egan sees the source text as a verbal prompt that is constant for all translators; thus, the source text serves as the *tertium comparationis*.

Åke Viberg's paper "What happens in translation" explores differences between original text and translations and the implications of such differences for contrastive studies by comparing the use of verbs meaning *sit*, *stand* and *lie* in original and translated texts from the English-Swedish Parallel Corpus (ESPC). Viberg shows that postural verbs have a much higher frequency overall in Swedish originals than in English originals, and also that postural verbs are significantly under-represented in the Swedish translations and significantly over-represented in the English translations. These findings point to contrastive differences between English and Swedish, as well as to translation effects. Importantly, however, Viberg shows that the pattern of over- and under-representation looks slightly different at a more fine-grained level of analysis: that of different types of subjects. Viberg discusses his findings both through the lens of research methodology and from a theoretical point of view. A conclusion of importance for studies based on small-scale parallel corpora is that although frequencies can be considerably skewed in translations, a language remains true to its system of basic semantic contrasts in professional translations. Viberg also points to how his results can be related to theories of language contact and studies of second language acquisition and bilingual development – truly a new theoretical pitch in corpus-based contrastive studies.

The second section, *Contrasting word meaning and use*, includes two papers where the meaning and functions of Swedish words are viewed through their English translations and sources in the English-Swedish Parallel Corpus. In the first of these papers, Karin Aijmer uses the ESPC to explore the Swedish modal particle *nog*. Because English does not have modal particles, and thus no obvious equivalent of *nog*, translation patterns are suggestive of the meanings and functions of the

particle. *Nog* is often omitted in translations, or added in translations from English to Swedish. Overt translations tend to convey modal meaning, with modal adverbs being the most frequent type of correspondence. In medial position, *nog* typically functions as a softener or downtoner. In initial position, however, *nog* can be stressed and carries a more contrastive meaning, for instance involving contradictory assumptions. The key feature of *nog* (and its correspondences) is that the speaker, having sufficient knowledge to judge that something is true, assumes responsibility for the truth of his/her utterances, attitudes and opinions, and actions. However, the modal meaning of *nog* serves as a mitigator of the opinion or attitude expressed.

Mats Johansson and Lene Nordrum take Bengt Altenberg's observations about the Swedish word *hinna* as their point of departure. Similarly to Karin Aijmer, they work from the assumption that since *hinna* lacks a straightforward equivalent in English, its English correspondences in the ESPC are particularly useful for shedding light on *hinna*'s meaning and use. Based on the observed translations and sources of *hinna*, Johansson and Nordrum propose that *hinna* is monosemous, but is usually enriched by implied meanings through presupposition and conversational implicature. The core-meaning of *hinna,* they suggest, is time sufficiency, but *hinna* typically presupposes ability and conversationally implicates actualization. The paper ends in the vein of Eckhard König's (2012) observation in a special issue of the journal *Languages in Contrast* regarding how results from small-scale contrastive investigations can serve as complements to general claims made in typology. In the case of *hinna*, its meaning categories can be fitted on van der Auwera and Plungian's (1998) semantic map of modality, in which case the meaning components time sufficiency and ability represent contiguous modal meanings.

together contribute to the pitch of the future key of corpus-based contrastive linguistics, at least in its Scandinavian accent.

References

Altenberg, Bengt. 1999. "Adverbial connectors in English and Swedish: Semantic and lexical correspondences". In Hilde Hasselgård & Signe Oksefjell (eds), *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi. 249–268.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.

Johansson, Stig. 2012. Cross-linguistic perspectives. In Merjä Kytö (ed.)*, English Corpus Linguistics: Crossing Paths*. Amsterdam: Rodopi. 45–68.

König, Ekkehard. 2012. "Contrastive Linguistics and Language Comparison". *Languages in Contrast*. 12(1): 3–26. DOI: 10.1075/lic.12.1.02kon

van der Auwera, Johan and Vladimir A. Plungian. 1998. "Modality's semantic map." *Linguistic Typology* 1998(2): 79-124.

Contrastive linguistics in a new key[1]

Jarle Ebeling, University of Oslo

In 1994 a group of researchers from the Nordic countries met in Lund, Sweden, for a symposium titled *Languages in Contrast: A Symposium on Text-based Cross-linguistic Studies*.[2] The purpose of the symposium was to gather researchers "active in the field of contrastive linguistics and who share an interest in linking cross-linguistic studies with corpus linguistics and the use of the computer as a tool in linguistic research" (Aijmer et al. (eds) 1996, Acknowledgments). Little did we know back then that we would meet again in Lund in 2014 to commemorate the symposium and celebrate the start of the collaborative efforts that led to the compilation of the parallel corpora known as the English-Norwegian Parallel Corpus (ENPC), the Swedish-English Parallel Corpus (ESPC) and the Finnish-English Contrastive Corpus Studies Project (FECCS).[3] Moreover, in 2014, we could also celebrate the 20th anniversary of the research network "Languages in contrast", which was generously funded by the Nordic Academy for Advanced Study (NorFa) between 1994-1996, and which meant that the parallel corpus projects could employ research assistants and a software developer (Knut Hofland) to work on the compilation of the corpora.[4]

---

[1] The phrase "contrastive linguistics in a new key" is borrowed from Johansson (2012: 46) who says of present-day contrastive linguistics "that the focus on immediate applications is toned down; the contrastive study is text-based rather than a comparison of systems in the abstract; and the study draws on electronic corpora and the use of computational tools".

[2] The only non-Nordic participant was Sylviane Granger from the Centre d'Études Anglaises, Université Catholique de Louvain. Sylviane Granger later became one of the compilers of the PLECI corpus, built on the same model as the ENPC/ESPC. See http://www.uclouvain.be/en-cecl-pleci.html.

[3] A Danish parallel corpus project was also proposed, but never materialised.

[4] Stig Johansson and Knut Hofland had already run experiments on the alignment of parallel texts in 1993 (Johansson & Hofland 1994), and the Swedish project "Text-based Contrastive Studies in English" had also started in 1993 (Aijmer et al. 1996).

Looking back, one may ask why the time seemed ripe for corpus-based contrastive studies based on parallel text corpora in 1994. For to be fair, parallel texts have existed almost since the dawn of writing, as evidenced by e.g. clay tablets with Sumerian and Akkadian aligned cuneiform texts from Assyria from the 2nd millennium BCE, let alone famous inscriptions such as the Rosetta Stone (see Véronis 2000 and Ebeling & Ebeling 2013). Furthermore, contrastive linguistics was not exactly a new discipline in 1994. Two years earlier, for instance, the 25th International Conference on Contrastive Linguistics and Cross Language Studies was held at Adam Mickiewicz University in Rydzyna, Poland, bearing evidence of the strong tradition of contrastive linguistics in the then so-called Eastern European countries (cf. e.g. Fisiak 1983 and Mair & Markus (eds) 1992). In fact, as early as 1970, a research group in former Yugoslavia translated parts of the Brown corpus and aligned the source (original) and target (translation) texts and built a parallel corpus (see e.g. Filipović (ed.) 1971). In the same decade, Jan Svartvik, one of the pioneers of corpus linguistics, initiated a project titled The Swedish-English Contrastive Studies project (Svartvik 1973) at the University of Lund.

Several important trends and developments in the late 1980s and early 1990s heralded the advent of parallel corpora and corpus-based contrastive analysis (cf. Hartmann 1997). In the late 1980s corpus linguistics was firmly established as a research paradigm, but had hitherto mostly been preoccupied with compiling monolingual corpora and building tools for exploring them. An irresistible challenge for a corpus linguist such as Stig Johansson, who had also been interested in contrastive linguistics since the 1970s (Johansson 1975), was to see if some of the techniques and tools developed for monolingual corpora could extend to bi- and multilingual corpora. Kari Sajavaara certainly thought so, but envisaged that it would take some time to develop these tools and techniques.

> It is no longer necessary for the contrastive linguist to invent examples. It is now possible to resort to corpora, where the relevant instances can be found by means of automatic searchers [reference omitted]. There is a wealth of information about principles to be applied in the compilation of unilingual text corpora [references omitted], but there is much less information about parallel corpora [references omitted]. Since a bilingual parallel corpus is different from a unilingual one, it is to be assumed that the principles for its compilation are also different. It is evident that

> within the next few years we will have more information about this problem. (Sajavaara 1996: 31)

About the same time as contrastive linguistics turned to corpora for new impetus, so did translation studies. In his 1995 book, Gideon Toury argues strongly for a descriptive, empirically based approach to translation studies.[5]

> [...], as I see it, one of the weaknesses of Translation Studies in the present phase of its evolution lies precisely in the fact that descriptivism as such is frowned upon, driving every other scholar to indulge in theorizing, very often in a highly speculative manner. (Toury 1995: 266)

The empirical basis for which Toury made a case came in the form of parallel and translation corpora of the kind developed by e.g. Mona Baker (Baker 1993).[6] A few years earlier, the translation scholar Brian Harris had written a short piece on what he called bi-texts, which he saw as a new concept in translation theory.[7] To Harris, a bi-text should be thought of not only as a complete source text (ST, original) and a target text (TT, its translation), but also as combinations of words and segments within the two texts, since

> translators do not translate whole texts at one fell swoop. They proceed a little at a time, and as they proceed each spurt, each segment forms a fragment of bi-text in their mind. Bi-text retains this structure when it is recorded on paper or in a computer: that is to say, not only is the whole text a bi-text but each segment combines ST and TT. (Harris 1988: 8)

These bi-texts could then be stored electronically and make up a hypertext base or translation memory, where the translator could search his or her own previous work. Such a hyper-bi-text system should

---

[5] In 1969 Hartmann writes: "I must repeat my plea for more empirical research into translation within the frameworks provided by applied and contrastive linguistics (Hartmann 1969 [2007: 24]).

[6] In July 1993, Mona Baker wrote to Stig Johansson saying that John Sinclair had showed her a copy of Stig's proposal for an English-Norwegian corpus, and that she hoped to be setting up her own corpus of translated texts soon at The University of Manchester Institute of Science and Technology (UMIST).

[7] Harris's ideas have also been reiterated in Véronis (2000) and Ebeling & Ebeling (2013).

include a search engine "programmed in such a way that when it finds an occurrence of the word [one is looking for], it retrieves and displays the whole translation unit in which it occurs" (ibid. 9).

Without specifying how, Harris foresees an even more sophisticated system whereby one can search for units similar to the one one is after, i.e. a system where what is combined is not translation units, but meaning units.

As to the display of bi-texts Harris advocates an interlinear translation display, where each line of the target text is interlaced between the corresponding line(s) of source text making up an interlinear bi-text. However, Harris (ibid. 11) adds that "[t]here remains the serious problem with interlinear bi-text, that of aligning ST and TT."

The serious problem mentioned by Harris was soon tackled by computer scientists interested in natural language processing.[8] In a series of experiments in the early 1990s Gale and Church showed how parallel texts could be automatically aligned at sentence level using sentence length in characters on text pairs already aligned at paragraph level (Gale & Church 1991, 1993). The method, though very simple, was surprisingly successful:

> The model was motivated by the observation that longer regions of text tend to have longer translations, and that shorter regions of text tend to have shorter translations. In particular, we found that the correlation between length of paragraph in characters and the length of its translation was extremely high (0.991). This high correlation suggests that length might be a strong clue for sentence alignment. (Gale & Church 1993: 89)

At the 14th ICAME conference in Zürich in 1993, Stig Johansson and Knut Hofland proposed a method for aligning English and Norwegian source (original) and target (translation) texts which can be seen as an amalgamation of several of the methods tried out in earlier experiments (see Johansson & Hofland 1994). Johansson and Hofland, though recognising the effectiveness of using sentence length in characters, wanted to explore a more linguistically-grounded method of aligning a pair of texts at sentence level and proposed to include an anchor word list as a central component. The anchor word list contains words and expressions "where the correspondence between the languages could be

---

[8] See Véronis (2000) for an overview of the many alignment methods explored and experiments performed at the time.

expected to be rather good" (Johansson & Hofland 1994: 30), and the object of using anchor words "was to calculate an *anchor score* which can be used in sentence alignment, expressing the number of shared anchor words" (ibid. 31).

The bilingual anchor word lists we use today contain function words, numerals, frequent and stable content words and names of days, months, countries and languages. In addition to sentence length in terms of characters and anchor words, the latest version of the alignment program, the Translation Corpus Aligner (TCA v.2) also takes special characters, e.g. %, ?, !, matching proper names and cognate words in the two languages into consideration when aligning sentences. Such bilingual word lists, of approx. 2,000 lines, have now been made for a range of language pairs.

In an investigation of the effect of the anchor word list carried out for the language pair English-Portuguese (Santos & Oksefjell 1999) it was shown that the anchor word list is essential to the success of the alignment. In fact, six out of the sixteen texts used in the investigation could not be aligned without the anchor word list, thus suggesting that a language-dependent aligner such as the Translation Corpus Aligner (Hofland & Johansson 1998) was an important step in the development of aligners for parallel texts.

As to the idea of performing contrastive analysis based on original and translated texts pitted against each other in such as fashion that similarities and differences become apparent, this owes a lot to James' (1980) belief in translation as the best *tertium comparationis*, Hartmann's (1980: 37-38) classification of types of parallel texts, and the method known as analytical comparison (e.g. Mathesius 1975), which is straightforwardly described by Čmejrková (1992: 5-6) in this way:

> As long as we knew only one language, it seemed to be the only one, writes the Czech typologist Vladimír Skalička [reference omitted]. The mother tongue seems to be the most natural, practical, and beautiful one. But awareness of foreign language solutions raises the awareness of the mother tongue and the speaker consequently asks questions he would have not asked otherwise.

We are now in a position to answer the question raised at the beginning of the second paragraph of this résumé, as to why the time was ripe for contrastive studies in a new key based on parallel corpora in 1994. It was, as is often the case, the coming together of people, ideas and technology. The people, in our context, were above all Stig Johansson,

Bengt Altenberg, Karin Aijmer and Knut Hofland.[9] As to the ideas, I have attempted to point to some of their sources, knowing quite well that great injustice has been done to many linguists and other scientists, who have had, and acted on, similar ideas over the years. Finally, technology: the corpora we began compiling in earnest in 1994 could hardly have been built without the technological advances that had taken place in the 1970s and 80s, with, e.g., the advent of the personal computer and the emerging new disciplines of computational linguistics and natural language processing.

Today we know that work on parallel corpora did not stop when ENPC/ESPC/FECCS were completed: several new copora have emerged, often including other language pairs than English and the Nordic languages. Moreover, corpus-based contrastive linguistics as a new discipline has gained momemtum with for example the pre-conference workshop as a fixture of ICAME conferences.

References

Aijmer, Karin, Bengt Altenberg and Mats Johansson (eds). 1996. *Papers from a Symposium on Text-based Cross-linguistic Studies. Lund 4-5 March 1994* [Lund Studies in English 88]. Lund: Lund University Press.

Aijmer, Karin, Bengt Altenberg and Mats Johansson. 1996. Text-based Contrastive Studies in English. Presentation of a Project. In Aijmer et al. (eds), 87–112.

Baker, Mona. 1993. Corpus Linguistics and Translation Studies: Implications and Applications. In Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds), *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins. 233–250.

Čmejrková, Světla. 1992. Linguistic Comparison and Linguistic Characterology. In Christian Mair and Manfred Markus (eds), 3-8.

Ebeling, Signe O. and Jarle Ebeling. 2013. From Babylon to Bergen. On the Usefulness of Aligned Texts. *Bergen Language and Linguistics Studies* (BELLS), 3:1. <DOI: http://dx.doi.org/10.15845/bells.v3i1>

---

[9] Mention should also be made of the many research assistants, including the current author, who worked hard on the various corpus projects, scanning, proof-reading and aligning texts.

Filipović, Rudolf. (ed.). 1971. *Zagreb Conference on English Contrastive Projects, 7-9 December 1970. Papers and Discussion.* Zagreb: Institute of Linguistics, Faculty of Philosophy, University of Zagreb. (Also *The Yugoslav Serbo-Croatian-English Contrastive Project, B. Studies 4.* Zagreb: Institute of Linguistics, Faculty of Philosophy, University of Zagreb.)

Fisiak, Jacek. 1983. Present Trends in Contrastive Linguistics. In Kari Sajavaara (ed.), *Cross-language Analysis and Second Language Acquisition* 1 [Jyväskylä Cross-languag Sudies 9]. Jyväskylä: University of Jyväskylä. 9–38.

Gale, William A. and Kenneth W. Church. 1991. A Program for Aligning Sentences in Bilingual Corpora. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics* (ACL), Berkeley. 177–184.

Gale, William A. and Kenneth W. Church. 1993. A Program for Aligning Sentences in Bilingual Corpora. *Computational Linguistics*, 19:3. 75–102.

Harris, Brian. 1988. Bitexts: A New Concept in Translation Theory. *Language Monthly* 54. 8–10.

Hartmann, Reinhard R.K. 1980. *Contrastive Textology*. [Studies in Descriptive Linguistics 5]. Heidelberg: Julius Groos Verlag.

Hartmann, Reinhard R.K. 1997. From Contrastive Textology to Parallel Text Corpora: Theory and Applications. In Hickey, R. and S. Puppel (eds), *Language History and Linguistic Modelling. A Festschrift for Jacek Fisiak on his 60th Birthday. Volume II* [Trends in Linguistics. Studies and Monographs 101]. Berlin: Mouton de Gruyter. 1973–1987.

Hartmann, Reinhard R.K. 2007. *Interlingual Lexicography. Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary*. Tübingen: Max Niemeyer Verlag. (Essays originally published between 1969 and 2005.)

Hofland, Knut and Stig Johansson. 1998. The Translation Corpus Aligner: A Program for Automatic Alignment of Parallel Texts. In Stig Johansson and Signe Oksefjell (eds), *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Amsterdam: Rodopi. 87–100.

James, Carl. 1980. *Contrastive Analysis*. London: Longman.

Johansson, Stig. 1975. *Papers in Contrastive Linguistics and Language Testing* [Lund Studies in English 50]. Lund: Liber-Läromedel/Gleerup.

Johansson, Stig. 2012. Cross-linguistic Perspectives. In Merjä Kytö (ed*.), English Corpus Linguistics: Crossing Paths*. Amsterdam: Rodopi. 45–68.

Johansson, Stig and Knut Hofland. 1994. Towards an English-Norwegian Parallel Corpus. In Udo Fries, Gunnel Tottie, and Peter Schneider (eds), *Creating and Using English Language Corpora*, Amsterdam: Rodopi. 25–37.

Mair, Christian and Manfred Markus (eds). 1992. *New Departures in Contrastive Linguistics. Proceedings of the Conference Held at the Leopold-Franzens-University of Innsbuck, Austria, 10-12 May 1991*. Volume I & II.[Innsbrucker Beiträge zur Kulturwissenschaft. Anglistische Reihe Band 4/5]. Innsbruck: AMOE.

Mathesius, Vilém. 1975. *A Functional Analysis of Present Day English on a General Linguistic Basis*. The Hague: Mouton. (Translation into English by Josef Vachek of "Obsahový rozbor současné angličtiny na základě obecně lingvistickém".)

Sajavaara, Kari. 1996. New Challenges for Contrastive Linguistics. In Aijmer et al. (eds), 17–36.

Santos, Diana and Signe Oksefjell. 1999. An Evaluation of the Translation Corpus Aligner, with Special Reference to the Language Pair English-Portuguese. In Torbjørn Nordgård (ed.), *NODALIDA'99, Proceedings from the 12th "Nordiske datalingvistikkdager"*. Trondheim: NTNU. 191–205.

Svartvik, Jan (ed.). 1973. *Errata. Papers in Error Analysis*. Lund: CWK Gleerup.

Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond* [Benjamins Translation Library 100]. Amsterdam: John Benjamins.

Véronis, Jean. 2000. From the Rosetta Stone to the Information Society: A Survey of Parallel Text Processing. In Jean Véronis (ed.), *Parallel Text Processing: Alignment and Use of Translation Corpora*. Dordrecht: Kluwer Academic Publishers. 1–24.

Such and *sådan* – the same but different

Johan van der Auwera, University of Antwerp, and Evie Coussé, University of Gothenburg

Abstract
English *such* has a variety of uses, which nearly always involve the expression of similarity, and *such* can therefore be called a 'similative'. Swedish *sådan* has very similar uses. However, the two similatives differ strongly with respect to the frequency of some of these uses. Thus *such*, different from *sådan*, rarely functions as a noun or pronoun, compared to *sådan* it more frequently exhibits the so-called 'intensifying use' and it has more competition from constructions with *like*, *sort* and *kind* (*sort*, *typ* in Swedish). The study uses three methods: (i) non-corpus analysis based on intuition, scrutiny of existing scholarship and google searches, (ii) analyis of comparable corpora, and (iii) analysis of parallel corpora.

1. Introduction

This paper aims to describe the similarities and the differences between English *such* and Swedish *sådan* and it does so using three different methodologies. The first is that of language-specific scholarship, relying on intuition, previous scholarship and web searches. This method gives us a good analysis of the possible uses of *such* and *sådan* and it tells us that these uses are very similar. The second method uses two language-specific corpora, and it shows which uses are more or less frequent and it tells us that *such* and *sådan* are very different. The third method uses parallel corpora and it discloses what other means the two languages have at their disposal to express *such* and *sådan*. The paper aims to advance the contrastive analysis of *such* and *sådan* and, at a more general level, it is a plea for methodological triangulation.

2. Such *and* sådan *are rather similar*

The English word *such* has attracted quite a bit of attention in recent years, in part because it has a great diversity of uses. Some of them are illustrated below. In (1) to (6) we focus on the categorial status of *such* and on word order; in (7) to (9) on meaning. Most examples are easily googlable quotes from famous people. Only for (7) and (9) do we give the URL.

(1)   Why was I born with *such* contemporaries? (O. Wilde)
(2)   An American has no sense of privacy … There is no *such* thing in the country. (G.B. Shaw)
(3)   If sex is *such* a natural thing, how come there are so many books on how to? (B. Midler)
(4)   There is no *such* thing as a non-political speech by a politician. (R. Nixon)
(5)   Before my term has ended, we shall have to test anew whether a nation *such* as ours can endure. (J. F. Kennedy)
(6)   Do not seek pleasure as *such*. (A. Lawrence Lowell)
(7)   We wanted to identify the communities with the best *such* opportunities in Mississippi. (http://www.nerdwallet.com/blog/reviews/best-towns-mississippi-young-families/, June 3 2015)
(8)   I had *such* a great time with the audience. (Lady Gaga)
(9)   If you say "so-and-so is vice president, finance, of *such and such*," should there be a comma after "finance"? (http://www.chicagomanualofstyle.org/qanda/data/faq/topics/Commas/faq0003.html, June 23 2015)

In (1) *such* could be an adjective or a determiner. In (2), however, *such* would seem to follow a determiner, so it can only be an adjective, but then in (3) *such* precedes a determiner and it cannot therefore qualify as either a determiner or an adjective. In (4) *such* is connected with *as*, but there is a noun in between *such* and *as* and this fact invites an analysis as a discontinuous construction. Yet in (5) *such* joins *as* in the postnominal domain. (6) shows another peculiarity: instead of *such (…) as*, it has *as such*. And if the examples (1) tot (5) give reason to think that *such* is indefinite, (7) shows that *such* has a definite use as well. In (8) the adnominal *such* has an 'intensifying' function (with *such a great* being equivalent to *a very great*) rather than an 'identifying' use – terminology due to Bolinger (1972). The strangest use of all is perhaps *such and such* in (9), which is non-compositional, i.e., it does not first refer to one type of thing and then to another. The diversity of uses has bewildered linguists. Referring to the research literature Wood (2002: 91) writes that "[t]he word *such* causes confusion and disagreement, and stipulations about its category often vacillate between hedges and contradictions" and fifteen years later van der Auwera & Sahoo (2015) argue that it is best to

treat *such* as a category all by itself, a one-member category, which they call 'similative', related, of course, to the determiner, the demonstrative and the adjective, but different from all.

The problem of categorizing *such* reappears with the translational equivalents in other languages, esp. Germanic ones: see e.g. Jensen (2013) on Danish, Fretheim (2010) on Norwegian, Demske (2005) on German, van der Auwera & Sahoo (2015) on Dutch, and Leu (2015: 161-165) on Germanic in general. For the Swedish similative *sådan*[1] there is Ekberg (2011a, b), who has pointed out that though *sådan* is normally preceded by the indefinite article *en* or *ett,* Malmö adolescents often leave out the indefinite article, in which case it does not have its normal meaning, translatable with 'such', but rather what has been called a 'recognitional' use (Himmelmann 1996). In this use, illustrated in (10), *sådan* introduces a new discourse referent, though of a type familiar to the hearer.

(10)   du   vet   jag   har   *sån*   säng
         you   know   I   have   such   bed
         "you know I have this bed"
         (Ekberg 2011b: 221-223)

Ekberg (2011a, b) also points out that, since the discourse referent of a recognitional uses is new, *sådan* comes close to doing the job of the indefinite article. The English similative does not have this use, only the demonstrative (notice *this* as a translation of *sådan* in (10)).[2] Another difference between Swedish and English is that with a full form, the one combining the similative and the indefinite article, English has the similative first (*such a*), but Swedish the indefinite article (*en sådan)*. So *such* and *sådan* are clearly different. Yet looking back at the illustrations for English, Swedish has counterparts for all of them, except for the postnominal *such as* use illustrated in (5) with *a nation such as ours*,

---

[1] In the text we will refer to '*sådan*', but it has six forms, all of the indefinite adjectival type, viz. *sådan, sådant, sådana* and their short forms *sån, sånt, såna*.
[2] Interestingly, the Dutch cognate to *such*, viz. *zulk,* does not allow a recognitional use either, but Dutch *zulk* is currently being replaced by the new similative *zo'n* (literally 'so a') and *zo'n* does have it (Van Olmen & van der Auwera 2014, Van Olmen 2015).

where Swedish either puts the similative in front of the noun or drops it altogether.

(11)   en   *sådan*   nation   som   vår …
       a   such   nation   as   ours

(12)   en   *nation*   (*sådan) som   vår ...
       a   nation   such   as   ours

Note that the non-decompositional *such and such* has a Swedish counterpart too, shown in (13).

(13)   Jag   heter   så   och   så,   jag
       I   am.called   so   and   so   I
       bor   där   och   där   med   de
       live   there   and   there   with   those
       och   de   familjemedlemmarna   och
       and   those   family members   and
       sådana   och   *sådana* husdjur ...
       such   and   such   pets
       "I am called so and so, and I live in such and such a place, with
       such and such family members and such and such pets and …"
       (http://amandaiunderlandet.ratata.fi/blogg/article-46556-44378-
       day-1-introduceyourself?offset_46556=825&goto=comments,
       June 23 2015)

Neither the Svenska Akademiens Grammatik (SAG) nor the Svenska Akademiens Ordbok (SAOB) mentions this use. Sådan also has the 'noun *as such*' construction and the definite use (both listed in *SAOB* but not in *SAG*[3]). (14) to (16) illustrate these uses.

---

[3] This definite use requires an adjective that implies uniqueness, both in English and Swedish, either a superlative or a word like *only/enda* and is probably typical for a formal, written register. *SAG* (III, 444) does draw attention to a definite use, claimed to be typical for spoken language, and this example is *Hennes såna buxor är mycket häftigare* "These pants of hers are much cooler".

(14) Det   är   rädsla   jag   känner   men   inte   för
      it   is   fear   I   feel   but   not   for
      situationen   som   *sådan*
      situation.the   as   such
      "It is fear I feel but not for the situation as such."
      (https://books.google.com/books?isbn=9137145037, June 23
      2015)

(15) Det   enda   *sånt*   besök   som   fungerade
      the   only   such   visit   that   worked
      för   oss   var   det   som   reseledaren
      for   us   was   it   that   guide.the
      styrde   upp.
      steared   up
      "The only visit that worked for us was one that the guide
       arranged."
      (http://www.mintur.se/forum/diskussion.php?save=1&forumDis
       cussionID=13522&page=1, September 25 2015)

(16) Under   det   tidiga   1990-talet   fann   man
      under   the   early   1990-number   found   one
      det   *första*   *sådana*   ämnet.
      the   first   such   substance.the
      "In the early 1990s one found the first such substance."
      (https://sv.wikipedia.org/wiki/Cannabinoid, June 23 2015)

Interestingly, in (15) *sådan* takes the 'normal' indefinite inflection, but
this is not quite felicitous, for the noun phrase is definite, and in (16)
sådan takes the definite inflection, which it is not supposed to have (*SAG*
II, 441). *Sådan* also allows the intensifying use, recognized in both *SAG*
and *SAOB* and illustrated in (17).

(17)   *Sådana*   fina   rosor!
       such   fine   roses
       "Such fine roses!"
       (*SAG* II, 447)

To conclude, *such* and *sådan* are rather similar.

3. Such *and* sådan *are rather different*

Even if *such* and *sådan* have very similar or identical uses, they do not avail themselves of these uses with the same frequency. The frequencies differ, both the overall frequency and the frequency for each use. We investigated the frequencies found in the original fiction parts of the English-Swedish Parallel Corpus (ESPC) (Altenberg et al. 2001). This corpus has the advantage of containing both original texts andtranslations, which will prove insightful in section 4. The figures do not, of course, give us a full picture, but it is at least a good start. The results for the overall frequencies are shown in Table 1.

Table 1. Overall frequencies of *such* and *sådan* (original English and Swedish texts ESPC)

	Absolute frequency	Normalized frequency
such	171 in 340,745 words	5.02 per 10,000 words
sådan	238 in 308,160 words	7.72 per 10,000 words

We thus see that Swedish *sådan* is more frequent than English *such* and this difference is statistically significant (χ^2 = 18.771, df = 1, p < 0.0001). The next question is whether this higher frequency is distributed over all its uses or rather due to just a few.

The main uses of both *such* and *sådan* are the adnominal uses, viz.uses (1) to (5) and (7), but not (6) and (9). For English the adnominal uses make up 89% of all uses, for Swedish quite a bit less, viz. 57%. The frequencies of all adnominal uses are shown in Table 2.

Table 2. Overall frequencies of adnominal uses of *such* and *sådan* (original English and Swedish texts ESPC)

	Absolute frequency	Normalized frequency
such	152 in 340,745 words	4.46 per 10,000 words
sådan	136 in 308,160 words	4.41 per 10,000 words

We see that the normalized numbers are very close and the difference is not, of course, statistically significant ($\chi^2 = 0.008$, df = 1, p = 0.93). This means that the higher frequency of *sådan* will be due to one or more of the non-adnominal uses. The uses in question are the nominal and pronominal uses.

(18) En   *sån*   skulle   man   haft   när   man   var
       a   such   should   one   had   when   one   was
       grabb. (SW1)[4]
       boy
       One should have had one of these when one was a boy. (SW1T)

(19) Varför   ska   vi   släpa   på   tunga trasmattor
       why   shall   we   drag   on   heavy rag.rugs
       när   det   finns   bra   *såna*   i
       when   it   is.found good   such   in
       plast? (SC1)
       plastic
       Why should we bring in heavy rag rugs when there are good ones
       in plastic? (SC1T)

(20) *Sånt*   som   ingen   kunnat   tänka   ut. (KE1)
       such   as   nobody   could   think   out
       Something that nobody could have come up with. (KE1T)

(21) Det   är   vidrigt   av   dig   att   säga   något
       it   is   disgusting of   you   to   say   something
       sådant. (MS1)
       such
       It is disgusting of you to say something like that. (MS1T)

(18) and (19) are unquestionably nominal, and (20) and (21) are arguably pronominal, though *SAG* (II, 444) would consider (20) and (21) to be nominal. The exact categorization is not relevant here and we will just

---

[4] The code between brackets refers to a unique text identifier in the ESPC (Altenberg et al. 2001:7). The text identifier in this example stands for the novel *Polismördaren* of Sjöwall & Wahlöö. The translated version has the same text identifier with the addition of the letter T.

refer to these uses as '(pro)nominal' ones. What is relevant is that English cannot translate *sådan* in (18) to (21) with *such*.

(22)   *One should have had a *such* when one was a boy.
(23)   *Why should we bring in heavy rag rugs when there are good *such* in plastic?
(24)   *Such* as nobody could have come up with.
(25)   *It is disgusting of you to say something *such*.

(Pro)nominal *such* does exist though. The *such* in the *such and such* construction of (9) is arguably a case in point. And here are a few more types.

(26)   But some/many *such* have yet to be reported.
(27)   But *such* have yet to be reported.
(28)   There you can buy bags, wallets, and *such*.

The next question is how frequent (pro)nominal *such* and *sådan* are. It turns out that (pro)nominal *such* is very infrequent: in the original fiction subcorpus of English, there is only one attestation of (pro)nominal *such*, viz. the one shown in (29).

(29)   […] Plato was embraced and absorbed and Aristotle was rediscovered and acclaimed "the Philosopher" by *such* as Aquinas. (JH1)

Swedish *sådan* is very different: in the Swedish subcorpus (pro)nominal *sådan* uses account for no less than 38 % of all uses. It will be remembered that 57 % of all *sådan* uses are adnominal, this leaves us with 43 % non-adnominal uses, and thus nearly all are (pro)nominal.

The strong contrast between (pro)nominal *such* and *sådan* makes sense. English adnominals do not normally 'go nominal': they remain adnominal and use the prop word *one*, as in *green ones*. Swedish adnominals go nominal without a prop word, as in *bra såna* (in (19)) or *gröna* in (30), referring, for example, to green apples.

(30) Ge   mig   gröna.
      give   me   green
      "Give me green ones."

However, what holds for adjectives need not hold for similatives. Thus Dutch adjectives nominalize without prop words, but the similatives rarely nominalize. When one chooses green apples, (31) is perfect, but when one chooses apples of one type rather than another one, (32) is barely acceptable.

(31) Geef   me   groene.
      give   me   green
      "Give me green ones."

(32) ?? Geef   me   zulke.
        give   me   such
        "Give me those ones."

The Dutch adjectives are more like the Swedish ones, but the similatives are more like the English ones.

    The prominence of (pro)nominal uses is no doubt the biggest difference between *such* and *sådan*. But there are more. In what follows we will focus on the difference between identifying vs. intensifying uses. As pointed out already, English and Swedish allow both uses. (33) and (34) illustrate identification, and (35) and (36) intensification.

(33) I want to have *such* a hat.

(34) Jag   vill   ha   en   *sådan*   hatt.
      I   want   have   a   such   hat
      "I want to have such a hat."

(35) *Such* a wonderful saw!

(36) En   *sådan*   fin   såg!
      a   such   fine   saw
      "Such a fine saw!"
      (*SAG* II, 445)

Of course, in some contexts it is not clear which reading is meant. Thus (37) is an example that can only be disambiguated in context, as is shown by Ghesquière & Van de Velde (2011: 786) the pressure could be of a certain type or of a high intensity. Swedish (38) is also vague.

(37)   There should be no need for *such* pressure.

(38)   Jag   hade   aldrig   tidigare sett   en   *sådan*
       I   had   never   before   seen   a   such
       avgrund   i   ett   par   ögon. (BL1)
       chasm   in   a   pair   eyes
       I had never before seen such a chasm in a pair of eyes. (BL1T)

In the English subcorpus the distribution of identifying vs. intensifying uses vs. those that are vague between the two is shown in Table 3.

Table 3. Identifying, intensifying and vague uses of adnominal *such* (original English texts ESPC)

Identifying	Intensifying	Vague	Total
101	43	8	152
66 %	28 %	5 %	100 %

When we come to Swedish, the facts are more complicated, for Swedish *sådan* also has the recognitional use, already mentioned for the speech of Malmö adolescent studied by Ekberg (2011a, b), but in our view it is not restricted to this register. (39) is anexample from the ESPC.

(39)   Det   var   en   *sån*   där   typiskt   skör
       it   was   a   such   there   typical   frail
       gammal   röst. (MG1)
       old   voice
       It was one of these typical frail old voices. (MG1T)

The distribution of identifying, intensifying, vague and recognitional uses of *sådan* is shown in Table 4.

Table 4. Identifying, intensifying, vague and recogntional uses of adnominal *sådan* (original Swedish texts ESPC)

Identifying	Intensifying	Vague	Recognitional	Total
100	21	5	10	136
74 %	15 %	4 %	7 %	100 %

We see that *sådan* intensifies less than *such* (15 % vs. 28 %) and that it *sådan* identifies more (74% vs. 66 %). Furthermore, for the full picture we should also include the (pro)nominal uses in Swedish, which is particularly interesting, because the (pro)nominal uses are all identifying. Table 5 tabulates the identifying uses for (ad/pro)nominal *sådan*.

Table 5. Identifying uses for (pro/ad)nominal *sådan* (original Swedish texts ESPC)

	Identifying		Total
Adnominal *sådan*	100	74 %	136
(Pro)nominal *sådan*	90	100 %	90
(Pro/ad)nominal *sådan*	190	84 %	226

The new figures are those in the last row and the conclusion is clear: *sådan* is more identifying than *such.* The next section will make clear why this is so.

4. *Alternatives for* such *and* sådan

If *sådan* is more dedicated to identification than to intensification, one wonders what other means Swedish has for intensification, more particularly, for exactly that type that *sådan* can express, but does less often than *such*. Here the translations in the Swedish-English Parallel Corpus come in handy. Let us have a look at the translational equivalents of adnominal intensifying *such* in Swedish.

Table 6. Intensifying adnominal *such* translated into Swedish (Swedish translated texts ESPC)

	sådan	*så*	Other	Total
such + noun	10	3	5	18
such + adjective + noun	5	16	4	25
Total	15	19	9	43
	35 %	44 %	21 %	100 %

When English combines intensifying *such* with a noun and without an adjective, in 3 out of 18 cases, Swedish chooses to intensify an adjective that approximates the meaning of the English noun, and then the intensification happens with *så*, not with *sådan*, as in (40).

(40)　You're *such* a baby. (FW1)

Du	är	*så*	barnslig. (FW1T)
you	are	so	childish

In the majority of the cases, however, Swedish too uses an intensifying *sådan.* However, when English combines the intensifying *such* with both a noun and an adjective, Swedish uses the intensifying *sådan* only in 5 out of the 25 cases, as illustrated in (41).

(41)　*Such* a beautiful deep blue they'd made it […] (DF1)

De	hade	lyckats	få	en	*sån*	fin	blå
they	had	managed	get	a	such	nice	blue

färg	på	dem […] (DF1T)
color	on	them

The winning strategy combines the adjective and the noun with an intensifying *så.* (42) and (43) are two illustrations.

(42)　He'd never seen *such* a beautiful woman up close, in the flesh. (ST1)

Han	hade	aldrig	sett	en	*så*	vacker	kvinna
he	had	never	seen	a	so	beautiful	woman

på	nära	håll,	inte	i	levande	livet. (ST1T)
on	close	hold	not	in	living	life

(43)   Then I noticed that in a corner, across from where they ate with
        such innocent relish, […], was the ghost of their son. (BO1)

Så	la	jag	märke till	att	i	ena
so	laid	I	notice to	that	in	one
hörnet,	mitt	emot	dem	där	de	satt
corner	mid	across	them	where	they	sat
och	åt	med	*så*	oskuldsfull		aptit,
and	ate	with	so	innocent		appetite
satt	deras	sons	ande, […]. (BO1T)			
sat	their	son's	ghost			

The Swedish alternative makes good sense: when the person seen in (42)
is a very beautiful woman, this is because she scores high on a scale from
less to more beautiful, not on a scale from less of a woman to more of a
woman. We can now also explain the puzzle left at the end of the
preceding section. *Sådan* is less intensifying than *such*, because when the
noun phrase has an adjective, Swedish typically intensifies with *så*.
Though this strategy is available in English too, it seems to be relatively
infrequent: in our corpus it did not occur at all. (44) is a made-up
example.[5]

(44)   He'd never seen *so* beautiful a woman up close, in the flesh.

We have also seen that *sådan* is more (pro)nominal than *such*. Table 7
shows how English translates (pro)nominal *sådan*.

Table 7. Identifying (pro)nominal *sådan* translated into English (English translated texts
ESPC)

such	*like/sort/kind*	Other	Zero	Total
9	39	36	6	90
10 %	43 %	40 %	7 %	100 %

Some translations use *such* and they are all adnominal (see (45) and (46)).

---

[5] An anonymous reviewer checked the relative frequency of the *such a +*
adjective + noun and *so* + adjective + *a* + noun strategies in the BNC, and found
the first one to be ten times as frequent as the second one.

(45)   Hur   går   *sådant*   till? (AP1)
       how   goes   such   to
       How do *such* things happen? (AP1T)

(46)   Det   är för   att   förleda   *sådana*   som   jag. (BL1)
       it   is for   to   mislead   such   as   I
       It is done to mislead *such* people as me. (BL1T)

(47) and (48) illustrate the 'Other' and 'Zero' categories.

(47)   Om   *sånt*   som   hör   fruntimmer   till […] (GT1)
       about   such   that   belongs   women   to
       About things that deal with females […]. (GT1T)

(48)   En   *sådan*   som   han   skulle   jag
       a   such   like   him   should   I
       säkert   ha   kommit ihåg […] (BL1)
       definitely   have   remembered
       But I'm sure I would have remembered him. […] (BL1T)

In (49)-(51) we illustrate a type where the notion of similarity is not expressed by *such* but by *like, sort of* or *kind of*. *Like* goes with a noun and *sort* and *kind* are nouns, at least, to some extent (though with *sorta* and *kinda* we see these constructions acquiring an adnominal status).

(49)   Jag   skämtar   aldrig   om   *sånt*   här. (SW1)
       I   joke   never   about   such   here
       I never kid about things like that. (SW1T)

(50)   *Sådant*   och   mycket   annat   berättar   min
       such   and   much   other   tells   my
       bror. (RJ1)
       brother
       That's the sort of story my brother tells me. (RJ1T)

(51) Ni   begriper   er   ju   på   *sånt*
     you   understand yourself   well   on   such
     här. (SW1)
     here
     You people from National Homicide understand this kind of case.
     (SW1T)

The *like/sort/kind* strategies are no less prominent for the translation of adnominal *such* (see (52) to (54)).

Table 8. Identifying adnominal *sådan* translated into English (English translated texts ESPC)

such	*like*/*sort*/*kind*	Other	Zero	Total
24	47	18	11	100

(52) En   *sådan*   dörr   kostar   nuförtiden   rätt
     a   such   door   costs   now   rather
     många   tusenlappar. (LG1)
     many   thousand.bills
     A door like that nowadays costs a pretty penny. (LG1T)

(53) Vi   är   inte   vana   vid   *såna*   saker. (SW1)
     we   are   not   used   with   such   things
     We're not used to this sort of thing. (SW1T)

(54) En   *sådan*   båt   är   det. (JMY1)
     a   such   boat   is   it
     That's the kind of boat it was. (JMY1T)

The other way round, Swedish has a *sort/typ* strategy too, but it cannot match the *sådan* stragegy.

Table 9. Identifying adnominal *such* translated into Swedish (Swedish translated texts ESPC)

sådan	*sort*/*typ*	Other	Zero	Total
61	10	12	18	101
60 %	10 %	12 %	18 %	100 %

(55)   […], among *such* people as she had never seen, […] (RDA1)
        […], bland   en   sorts   människor   som   hon   aldrig
        among   a   sort   people   who   she   never
        förr   hade   mött, […] (RDA1T)
        before   had   met

We can thus say that that similative *such* has more competition from other similative markers than Swedish *sådan*.

5. Conclusion

This paper is a demonstration of a three-pronged analysis of the differences and similarities between English *such* and Swedish *sådan.* Methodologically, we engaged in scholarship the way it is done outside the field of corpus linguistics, by essentially relying on intuition, by comparing – and criticizing – the existing language-specific analyses of *such* and *sådan*, and, in this day and age, web googling. We then added strictly corpus-based accounts of *such* and *sådan*, studying their uses both in original texts and in translatons. This triangulation allowed us to make clear that the possible uses of such and *sådan* are very similar, but their actual uses much less so. They both express similarity and we have called them 'similative', but English *such* is nearly only used adnominally, whereas Swedish *sådan* has prominent (pro)nominal uses as well. Another striking difference is that both have identifying and intensifying uses, the latter is more prominent for *such* than for *sådan*, and when the noun phrase contains an adjective, Swedish prefers to intensify with *så* rather than with *sådan.* Also, only in Swedish do we find the recognitional use. Finally, we have seen that to express similarity *such* has more competition from *like/sort/kind* strategies than *sådan* from *sort/typ* strategies.   Concluding the conclusion: the *such* and *sådan* similatives are neither the same nor different: they are … similar.

References
Altenberg, Bengt, Karin Aijmer & Mikael Svensson. 2001. "The English-
    Swedish Parallel Corpus. Manual of the enlarged version."
    http://www.sol.lu.se/engelska/corpus/corpus/espc.html

BNC = The British National Corpus, version 3. 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk/

Bolinger, Dwight. 1972. *Degree words.* The Hague: Mouton.

Demske, Ulrike. 2005. "Weshalb Demonstrativpronomina nicht immer Determinantien sind." *Deutsche Syntax: Empirie und Theorie*, Ed. Franz Josef d'Avis. Göteborgs Universitet. 53-80.

Ekberg, Lena. 2011a. "Extended uses of *sån* ('such') among adolescents in multilingual Malmö, Sweden." *Young Urban Swedish. Variation and Change in Multilingual Settings*. Eds. Roger Källström & Inger Lindberg. Gothenburg: University of Gothenburg. 49-64.

Ekberg, Lena. 2011b. "Joint attention and cooperation in the Swedish of adolescents in multilingual settings – The use of sän 'such' and såhär 'like'." *Ethnic Styles of Speaking in European Metropolitan Areas*. Eds. Friederike Kern & Margret Selting. Amsterdam: Benjamins. 217-237.

Fretheim, Thorstein. 2010. "Demonstratives and reference to individuated objects vs. reference to properties of objects: a contrastive analysis." *The Role of Data at the Semantics-pragmatics Interface.* Eds. Németh, Enikő T. & Karoby Bibok. Berlin: de Gruyter 95-124.

Ghesquière, Lobke & Freek Van de Velde. 2011. "A corpus-based account of the development of English *such* and Dutch *zulk*: Intensification and (inter)subjectification." *Cognitive Linguistics* 22: 765-797.

Himmelmann, Nikolaus P. 1996. "Demonstratives in narrative discourse: A taxonomy of universal uses." *Studies in Anaphora*. Ed. Barbara Fox. Amsterdam: Benjamins. 205-254.

Jensen, Eva Skafte. 2013. "Ordklasseproblemer: tilfældet *sådan*", *LexicoNordica* 20: 55-74.

Leu, Thomas. 2015. *The Architecture of Determiners*. Oxford: Oxford University Press.

[SAG] Teleman, Ulf, Staffan Hellberg, Erik Andersson. 1999. *Svenska akademiens Grammatik.* Stockholm: Norstedt.

[SAOB] Svenska Akademiens Ordbok. http://g3.spraakdata.gu.se/saob/

van der Auwera, Johan & Kalyanamalini Sahoo. 2015. "On comparative concepts and descriptive categories, such as they are", *Acta Linguistica Hafniensia* 47: 136-173.

Van Olmen, Daniel. 2015. "A corpus-based study of the evolution of Dutch *zo'n* 'so a'", Paper presented at the annual meeting of the *Societas Linguistica Europaea*

Van Olmen, Daniel & Johan van der Auwera. 2014. "Over *zo'n* en zo meer." *Patroon en argument. Een dubbelfeestalbum bij het emeritaat van William Van Belle en Joop van der Horst.* Eds. Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge. Leuven: Leuven University Press. 215-228.

Wood, Johanna L. 2002. "Much about such." *Studia Linguistica* 56: 91-117.

Does corpus size matter? Revisiting ENPC case studies with an extended version of the corpus

Signe Oksefjell Ebeling, University of Oslo

Abstract
The validity of contrastive findings that base themselves on material from small parallel corpora may be questioned, and ever since the compilation of the English-Norwegian Parallel Corpus (ENPC) and English-Swedish Parallel Corpus (ESPC) some 20 years ago we have been aware of this. Recently, the ENPC has been expanded into the ENPC+, holding bidirectional translation data three times the size of the fiction part of the original ENPC. Drawing on material from the ENPC+, this paper replicates three contrastive studies made on the basis of the fiction part of the original ENPC to explore to what extent corpus size matters. The replica studies suggest that individual style, genre and date of publication are variables that may have a greater impact on the results than mere corpus size.

1. Introduction

Parallel corpora are generally small in size and the validity of contrastive findings based on these corpora may be questioned as a consequence. Size is here understood in terms of the number of tokens, or running words, making up the corpora. Being aware of the relatively moderate size of parallel corpora, including the English-Norwegian Parallel Corpus (ENPC) and English-Swedish Parallel Corpus (ESPC),[1] researchers have been on the cautious side when making use of these corpora. Indeed, warnings and comments such as the following have commonly been expressed ever since the compilation of these corpora in the mid-1990s:

- Due to its restricted size, the corpus is not suitable for studies of collocations and lexical studies beyond the core vocabulary. (S. Johansson 1998a: 11)
- […] occurrences are too few to allow any generalisations. (S. Johansson 2008: 111)

---

[1] ENPC: http://www.hf.uio.no/ilos/english/services/omc/enpc/; ESPC: http://sprak.gu.se/english/research/research-activities/corpus-linguistics/corpora-at-the-dll/espc

- We need bigger corpora […] (S. Johansson 2009: 37)
- The number of examples is too small to be statistically reliable […] (M. Johansson 1996: 135)
- Provided that the material is large enough, MC values are thus a useful means of establishing semantic paradigms […] (Altenberg 1999: 266)
- Admittedly, the unit is not a very frequent one in the present data [...] (Ebeling et al. 2013: 191)

And finally, although mentioning the limited size of parallel corpora, Viberg (2010) has a more positive outlook on the matter:

- The limited size of this and several other parallel corpora is a temporary problem. (Viberg 2010: Section 2.1)

With regard to the first bullet point, it should be noted that contrastivists working with small-size parallel corpora have paid heed to size and have largely focused on high-frequency words, constructions or categories. This is in line with the commonly held view that "optimum corpus size depends on the specific linguistic investigation to be undertaken" (Granger 1998: 11, with reference to de Haan 1992).

The aim of this paper is to investigate to what extent the almost apologetic tone in the quotations above was justified when presenting some of the findings from the original parallel corpora, constantly drawing attention to the limited material at hand. With a bidirectional corpus three times the size of the fiction part of the original ENPC, viz. the ENPC+, the current paper will revisit a few case studies and compare findings based on the smaller (original) version of the corpus with those of the expanded version. The following lexis-based case studies will be considered:

- Ebeling (2003) on the Norwegian pseudo-coordination construction *bli* + present participle + *og* + infinitive;
- Johansson (1998b) on loving and hating in English and Norwegian;
- Johansson & Løken (1997) / Johansson (1998a) on some Norwegian discourse particles and their English correspondences.

The article is structured as follows: Section 2 begins with a description and comparison of the two versions of the ENPC. In Section 3, each of

the original case studies is introduced in turn, and the findings are compared with those of the ENPC+ follow-up studies. Potential implications of the overall findings will be discussed in Section 4, while Section 5 offers some concluding remarks and future prospects.

2. The ENPC vs. the ENPC+

The structure and the contents of the ENPC have previously been described in several publications (e.g. Johansson & Hofland 1994, Johansson 1998a, Oksefjell 1999), and only a brief outline will be offered here. The corpus is parallel in the sense that it contains comparable fictional and non-fiction texts in English and Norwegian as well as translations of the texts from and into the two languages. This parallelism, and bidirectionality, is captured in the oft-repeated illustration in Figure 1.
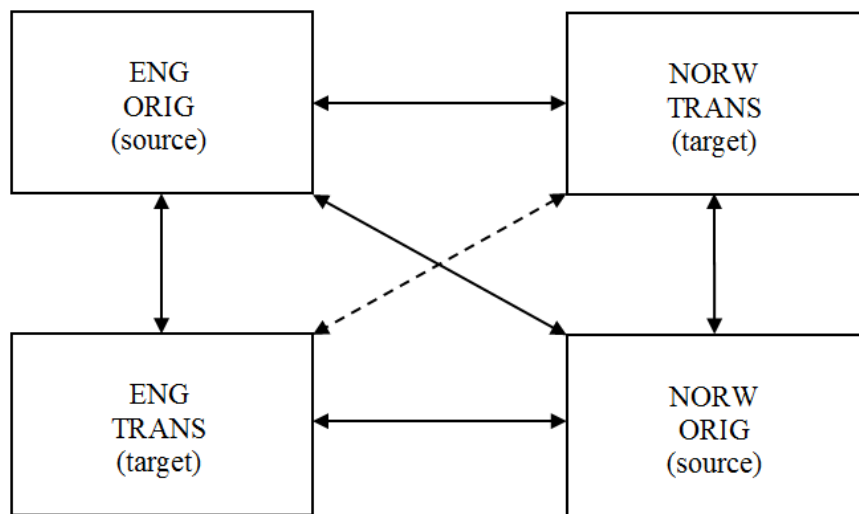

Figure 1. The structure of the ENPC (Johansson & Hofland 1994: 26)

It should be pointed out that the present investigation is concerned with the fiction part of the ENPC, as the expansion only includes fictional texts. Henceforth the original ENPC will be referred to as ENPCfiction.

   The expansion of ENPCfiction, resulting in the ENPC+, was mainly prepared in 2011-2012, almost 20 years after the initial stages of the

compilation of the ENPC (see Johansson & Hofland 1994). A full description of the ENPC+ and a comparison between ENPCfiction and its expansion are also offered elsewhere (Ebeling & Ebeling 2013: 86ff), but some of the main points are repeated here in Table 1.

Table 1. Main differences between ENPCfiction and its expansion (cf. Ebeling & Ebeling 2013: 87)

ENPCfiction	Expansion
Children's fiction, detective fiction and general fiction	Mainly crime (detective) fiction, some general fiction
Varieties of English (e.g. American, British, Canadian, South African)	Mainly British English
Texts (mainly) from the 1980s and 1990s	Texts from 2000 to 2012
Varieties of Norwegian (*bokmål* and *nynorsk*)	Norwegian *bokmål*
55 writers	13 writers
47 translators	10 translators
Extracts of 10,000–12,000 words	Complete books
400,000 x 2 words of original fiction material + their translations	900,000 x 2 words of original fiction material + their translations

With reference to Table 1, it is easy to point to variables that may have an impact on the comparison of studies based on ENPCfiction and the ENPC+, and it is not unproblematic to lump the two sets of data into one corpus. First of all, the fact that the expansion contains more text distributed across fewer writers and translators opens up for more idiosyncracies due to individual variation. Second, the texts contained in the expansion are naturally of a more recent date; thus diachronic change will have to be considered as a possible variable. Finally, variety and genre may also play a role, and perhaps genre in particular can skew the findings of the current investigation in certain directions. These potential pitfalls notwithstanding, the two parts have been brought together in the ENPC+.

An important issue in corpus linguistics in general, and another conceivable problem in the current study, is the degree of replicability of the studies. Although all the studies under scrutiny give relatively detailed information on how the actual linguistic classification was carried out, a 100% match between the original case study and the replica study cannot be guaranteed. Interestingly enough, this is also true of the study I performed myself some 12 years ago (Ebeling 2003),

serving as a reminder of the importance of including precise descriptions of the criteria used in the classification of corpus data.

3. Studies revisited

The following sections focus on three contrastive topics previously studied on the basis of the original version of ENPCfiction. The studies chosen to be replicated are all well-known to me; one is self-authored and the others are, or have been, part of my taught syllabus.

Each original study will be introduced briefly before the ENPC+ material is added for comparison. First, Section 3.1 concerns itself with Ebeling's (2003) investigation of Norwegian *bli* 'remain' as part of pseudo-coordination and its English correspondences.[2] Section 3.2 takes a closer look at Johansson's (1998b) comparison of the English verbs *love/hate* and their Norwegian counterparts *elske/hate*. Finally, Section 3.3 deals with Norwegian discourse particles and their English correspondences, using Johansson & Løken (1997) as a starting point, but mainly drawing on Johansson (1998a), who shifts the focus of attention to *probably*, one of the English correspondences of the Norwegian discourse particle *nok*.

3.1 Ebeling (2003)

The Norwegian sequence *bli* 'remain' + present participle + *og* 'and' + infinitive is a type of pseudo-coordination, i.e. it is a construction expressing hypotaxis rather than parataxis, even if the coordinating conjunction *og* is present (Vannebo 1969). Pseudo-coordination in Norwegian mainly consists of a posture verb + coordinated verb, as shown in example (1). In addition, the combination dealt with here has *bli* 'remain' as an auxiliary verb.

(1)   Han <u>blir stående og se</u> seg om, forundret, for alle er helt stille, og
        det er ikke Pappen akkurat vant til. [LSC1]
        Lit.: He remains standing and look around …

---

[2] Note that *bli* is a highly polysemous and versatile verb corresponding to a variety of English high-frequency verbs – notably *be*, *become*, *get* and *remain* – depending on context (see further Ebeling 2003).

This study was part of a larger contrastive investigation of the two Norwegian verbs *bli* and *få* and their correspondences in English; my focus in the sub-study of *bli* as part of pseudo-coordination was inevitably on what happened to *bli* in translation between English and Norwegian.

The distribution of types of English translations is given in Table 2, where it is shown that a so-called "synthetic" translation is found in almost 80% of the cases, as shown in the column in greyscale (ENPCfiction). By synthetic is meant cases where *bli* + present participle are merged into one verb, i.e. the one corresponding to the present participle, as in examples (2) and (3).

(2)   Han <u>blir stående</u> og se seg om, forundret, for alle er helt stille, og
       det er ikke Pappen akkurat vant til. [LSC1]
       He <u>stands</u> there, looking around, surprised because everyone is
       completely still, and Woody is not exactly used to that. [LSC1T]

(3)   Jeg <u>blir stående</u> og lytte. [ToEg1N]
       I <u>stand</u> listening to it for a while. [ToEg1TE]

Table 2. English translations of *bli* when *bli* is followed by a present participle + *og* + infinitive in ENPCfiction vs. ENPC+[3]

	ENPCfiction	ENPC+
translation	no.	no.
keep	1	5
be	2	4
remain	4	8
Ø	4	19
'synthetic'	43 (79.6%)	114 (76%)
Total	54	150

As shown in Table 2, other, marginal, translation types include *keep*, *be*, *remain* and Ø. An example of *remain* as a translation of *bli* is given in (4).

(4)   Midt på gulvet <u>ble</u> hun <u>stående</u> og se seg om en god stund, og jeg
       ventet engstelig. [EHA1]

---

[3] The ENPCfiction part of Table 2 is taken from Ebeling (2003: 169).

> In the center of the room she <u>remained standing</u>, looking around
> for a few moments while I waited anxiously. [EHA1T]

54 occurrences of pseudo-coordination with *bli* were recorded in
ENPCfiction. In the ENPC+ this is almost trebled to 150 occurrences, as
shown in Table 2. Given that the ENPC+ is three times the size of
ENPCfiction, this corresponds to the expected increase. In addition, a
fairly similar distribution of translation types is found in the ENPC+
study. Thus, the relationship between the Norwegian construction and its
English translation patters appears to be stable.

The picture emerging from this overview suggests that, in the case of
this particular construction, similar conclusions can be drawn on the
basis of less data. One of the conclusions of the original study at this
stage of the discussion was that there is no clear English counterpart of
bli as part of this construction, although a few examples do occur, as
with *remain* in example (4) above.

The main finding, however, was that the auxiliary *bli* tends to be
absorbed in the posture verb in translations into English, which tries to
capture the continuative nature of Norwegian *bli* + present participle.
This is particularly evident in instances such as (2), where the adverb
there has been added in the translation, seemingly to get the durative
element of the Norwegian construction more clearly across in the English
translation (see also Ebeling 2015b). A similar conclusion can be drawn
on the basis of the ENPC+ material; thus, size does not seem to matter in
this case.

3.2 Johansson (1998b)[4]

Before I introduce the original study by Johansson, it should be
mentioned that this particular study has also been revisited in a couple of
other papers, albeit with a different focus: Hasselgård (2011) in a
contrastive study of spoken English and Norwegian and Ebeling (2015a)
in a contrastive study of written English and Portuguese. Both of these
offer interesting similarities with, and additions to, the present

---

[4] A slightly revised version of the article is published as Chapter 5 in Johansson
(2007).

investigation, providing broader insights into the discussion of love and hate verbs across languages.

Johansson's study entitled "Loving and hating in English and Norwegian" was triggered by a couple of sentences appearing in Norwegian newspapers; one of which is repeated here as example (5). Johansson established that this news item was a direct translation from English, and the original version is offered below the Norwegian example (cf. Johansson 1998b: 93).

(5)   Jeg hater å bringe sladderen videre.
        (The original version: 'I hate to pass gossip on' was attributed to
        Shirley MacLaine and quoted in *Østlandets Blad*)

Johansson's immediate reaction was that the use of Norwegian *hate* in this context did not ring quite idiomatic, and he decided to examine the relationship between English *hate* and Norwegian *hate*, and also added their more loveable opposites: *love* and *elske*. More specifically, he investigates the relationship between these verbs in English and Norwegian, with particular attention to the types of objects they typically occur with. He applies the broad categories of *personal* and *non-personal* object. The distribution, referred to as percentages in the original study, is found in Table 3.[5]

Table 3. Distribution of objects with *hate, elske, hate* and *love* (the ENPCfiction distribution from Johansson 1998b: 95)

	ENPCfiction		ENPC+	
	Personal objects	Non-personal objects	Personal objects	Non-personal objects
N hate	65%	35%	54%	46%
N elske	61%	39%	59%	41%
E hate	27%	73%	19%	81%
E love	46%	54%	49%	51%

Johansson notes that in Norwegian, "the verbs take a personal object in the majority of cases, while non-personal objects are more common in English original texts" (p. 95). Examples include (6), where Norwegian

---

[5] Johansson also looks into the use of these verbs in the translated texts and notes some discrepancies between originals and translations. This study, however, will limit itself to the original texts only.

elske is followed by a personal object, and (7), where English *hate* is followed by a non-personal object in the form of a non-finite clause.

(6) Jeg trodde henne når hun stadig fortalte at hun <u>elsket meg</u>. [JW1]
     I believed her when she constantly told me that she <u>loved me</u>. [JW1T]

(7) He did once tell me that he <u>hated shaking hands</u>. [RDA1]

As can be seen in the ENPC+ columns in Table 3, this is still the case, but with some notable differences with regard to proportion. While the numbers for *elske* and *love* are more or less unchanged, the situation for *hate* and *hate* is different, as there is a marked increase of non-personal objects in both languages in going from ENPCfiction to the ENPC+. More than anything, and as the overall frequency of these verbs does not deviate much from the overall expected increase (i.e. the frequency is expected to treble, see Table 4), this seems to point to a change in use of the two verbs, rather than being a matter of having more data at our disposal.

Table 4. Comparison of total distribution of the verbs in ENPCfiction vs. ENPC+ (the ENPCfiction figures are taken from Johansson 1998b: 95)

	ENPCfiction	ENPC+
	Original texts	Original texts
N hate[6]	23	59
N elske[7]	36	122
E hate	67	212
E love	100	303

This potential language change is substantiated by the fact that Johansson found no occurrences in his material of *hate* + clausal complement in the Norwegian original texts, while there, in the ENPC+ material, are three such instances, one of which is found in example (8).

---

[6] The form *hata* 'hated' seems to have been excluded from Johansson's study; thus, this form has also been left out here (six occurrences in the material altogether).

[7] The form *elska* 'love/loved' seems to have been excluded from Johansson's study; thus, this form has also been left out here (seven occurrences in the material altogether).

(8)   Men hun <u>hatet å jogge</u>. [JoNe2N]
      But she <u>hated jogging</u>. [JoNe2TE]

This is precisely the kind of complementation that triggered the original study; Johansson states that his "immediate reaction was that these were anglicisms" (p. 93). My guess is that he was right; moreover, the more recent data also suggest that this pattern is on the increase in Norwegian. Johansson notes that, "[c]hanges of this kind are natural wherever there are languages in contact, but it is important to be aware of what is going on" (p. 102). The fact that such a change has taken place, or is taking place as we speak, was also confirmed by Hasselgård (2011) who performed an English-Norwegian contrastive study of spoken material of a more recent date than the written material in ENPCfiction.

Furthermore, it seems to have become more common, both in English and Norwegian to hate non-personal objects, thus suggesting that the force of the verbs may have been weakened. As pointed out by Johansson (p. 101):

> Whereas Norwegian *hate* and *elske* express a strong feeling and typically with a personal object, English *hate* and *love* are also used in a weakened sense [...]. The weakened sense is most likely to appear where the verbs combine with non-personal objects, particularly complement clauses.

This tendency of achieving a more weakened sense seems to be on the increase in both languages as far as *hate* is concerned; however, the three instances of complement clauses following N *hate* notwithstanding, the increase seems to be more prominent in non-personal NP complementation, as the percentage of complement clauses seems to be fairly stable (in Johansson's material, 26.5% of the non-personal objects of E *hate* were complement clauses, while the percentage in the ENPC+ material is 27.2%).[8]

Johansson himself draws attention to the question of corpus size, and says: "Judging by our limited material, it seems as if *elske* is more compatible with a following infinitive than *hate*" (Johansson 1998b: 99). This seems only to be marginally the case in the ENPC+ material, where 12% of the non-personal objects of *hate* and 13% of the non-personal

---

[8] In the ENPC 13 out of 49 non-personal objects were complement clauses; in the ENPC+ 34 out of 125 were complement clauses.

objects of *elske* are infinitive clauses. Thus, I will maintain that, rather than increased size, it is the more recent text material of the ENPC+ that explains the difference in non-personal objects with *hate*/*hate*; Norwegian in particular seems to be undergoing a change in accepting these complementation patterns more readily than was the case 20-30 years ago.

3.3 Johansson & Løken (1997) / Johansson (1998a)

Pre-dating the completion of the ENPC, these two studies were based on a slightly smaller sample than ENPCfiction. More specifically, the sample contains 27 fiction texts in each direction of translation instead of 30 (approx. 360,000 words in each part of the corpus, amounting to roughly 1.4 million words in total compared to the ENPC+ with 5.2 million; i.e. the ENPC+ is 3.7 times larger than the version of the ENPC used in the two studies. This will of course be taken into consideration in the comparison below.

One of the discourse particles Johansson & Løken deal with in some detail is *nok* and its correspondences in English. Table 5 gives an overview of the distribution of these correspondences – both translations and sources – in the ENPCfiction material.

Table 5. Correspondences of the Norwegian modal particle *nok*, expressed in percent within each column (cf. Johansson & Løken 1997: 168-169; Johansson 1998a: 14)

	ENPCfiction	
Correspondence	E translations (N = 141)	E source (N = 79)
probably	25	6
other adverb	21	4
verb construction	11	10
clause	9	10
miscellaneous	3	5
zero	31	65

As mentioned in Section 3, the study by Johansson & Løken (1997) on Norwegian discourse particles and their English correspondences served as a starting point for Johansson (1998a). While the discussion of *nok* in Johansson & Løken (1997) is concerned with all of its English correspondences and the lack of a clear English counterpart, as evidenced by the strikingly high number of zero correspondences,

Johansson (1998a) continues the discussion by looking at it from the reverse perspective, focusing on the main single overt translation correspondence *probably*.

He starts by mapping the Norwegian correspondences of *probably*, and notes that the frequency of zero correspondences is low and Norwegian modal particles are infrequent as translations of *probably* (see Table 6). "A plausible interpretation of these results", he says, "is that the existence of close formal and semantic correspondences simplifies the task of the Norwegian translator. By contrast, when faced with the problems of rendering Norwegian *nok*, the English translator finds no easy solution" (p. 15).

Table 6. Correspondences of the English adverb *probably*, expressed in per cent within each column (cf. Johansson 1998a: 15)

Correspondence	ENPCfiction	
	N translation (N = 94)	N source (N = 141)
nok	3	25
vel	6	28
antagelig(vis) [antakelig]	21	3
kanskje	3	9
sannsynligvis	37	16
sikkert	11	9
trolig	3	1
miscellaneous	13	6
zero	2	4

The concern in the present investigation is whether the same conclusions can be drawn when the study is based on material from a more sizeable corpus. The replica study will focus on *probably* and its translations into Norwegian only.

Comparing the number of occurrences in the two versions of the corpus (see Table 7), we can observe a striking difference in the frequency with which *probably* occurs: 94 vs. 580, which means that it is more than six times as frequent in the ENPC+ compared to ENPCfiction. Given that the ENPC+ is 3.7 times larger than the version of the ENPC that Johansson used, the expected frequency of *probably* in the ENPC+ would be around 350. Another striking observation that can be made from Table 7 is the use of *antagelig, antakelig* and *antageligvis,* for which the most frequent variant – *antakelig* – is used as shorthand for all

three, in the more recent material. The *antakelig*-words in Johansson's material accounted for 21% of the occurrences, while they account for almost 50% in the ENPC+ material. This increase seems to have taken place at the expense of *sannsynligvis* in particular (but also other items in the miscellaneous category).

Table 7. Translations of the English adverb *probably*, expressed in per cent within each column, in ENPCfiction vs. ENPC+

	ENPCfiction	ENPC+
Correspondence	N translations (N = 94)	N translations (N = 580)
nok	3	5.1
vel	6	7.4
antakelig	21	47.7
kanskje	3	1.6
sannsynligvis	37	10.7
sikkert	11	16.5
trolig	3	1.9
miscellaneous	13	6.4
zero	2	2.7

In the study on *love* and *hate* (Section 3.2), linguistic change as a result of the time span between the two parts of the corpus seemed to account for some of the differences that were noted between the two versions of the corpus. In the current context, however, it is less likely that the increased use of *probably* is due to linguistic change. In fact, there rather seem to be two writers (and three texts) in particular that contribute towards the increased use of *probably* in the ENPC+, namely PeRo1E (90 occ.), PeRo2E (102 occ.) and TaFr1E (114 occ.).[9]

Admittedly, the word count for these texts is around ten times higher than the word count for the text extracts of ENPCfiction; nevertheless, the use of *probably* is proportionally much higher in those three texts.

In some sense, the three texts referred to can be characterised as what Sinclair calls "rogue texts":

> In any variety of a language there will be some texts — "rogue" texts — which stand out as radically different from the others in their putative category. (Sinclair 2005: 13)

---

[9] PeRo = Peter Robinson; TaFr = Tana French. See Ebeling & Ebeling (2013: 241ff) for a full overview of authors and texts included in the ENPC+.

While the texts may not be so different from the others in their category overall, they are radically different in the use of particular words. In the light of Sinclair's observation regarding "rogue texts", an experiment of excluding the three texts in question was conducted. Thus a version of the ENPC+ 2.3 times larger than the version of ENPCfiction used by Johansson was produced.

When excluding the three texts in question, it can be seen, in Table 8, that the number of occurrences of *probably* increases by 2.9 compared to ENPCfiction (instead of 6.2, which was the case above), which is much closer to the expected increase of 2.3. A log-likelihood test shows that the difference is not statistically significant.

Table 8. Number of occurrences of *probably* in the reduced version of the ENPC+ vs. ENPCfiction

ENPCfiction	ENPC+ (excluding PeRo1E, PeRo2E, TaFr1E = 835,000 words = 2.3 times the size of Johansson's original sample)	
(N = 94)	(N =274)	LL = 3.79; p > 0.05 (difference is not statistically significant)

However, while *probably* is now seen to increase by roughly what would be expected, the translation correspondences still show a bias in the use of *antakelig* at the expense of *sannsynligvis*, Although one translator in particular seems to favour *antakelig*, this does not fully explain what is going on – whether *antakelig* is generally on the increase at the expense of *sannsynligvis*, or whether it is only tied to the preferences of the individual writers and translators represented in the corpus.

Such observations are of course far from new: corpus linguists have always been aware of individual variation and idiosyncracies on the part of the writers (and translators), and it seems to be a more pertinent problem in the use of individual words than in the use of grammatical patterns. Individual lexical items appear to stand a greater risk of turning into some sort of pet word. In the future, proper statistical measures on these issues should be carried out, but it is beyond the scope of this paper.

In connection with the use of words such as *probably* and *antakelig* it is also important to mention that they are often used in dialogue, and dialogue is a feature that, according to de Haan (1996), is more characteristic of crime fiction than general fiction.

> It is obvious that the three crime texts contain far more dialogue than any of the other four texts (roughly 60 per cent of the total number of sentences in each of the three crime texts). (de Haan 1996: 26)

The three texts that were excluded are all in the category "crime fiction", which means that they are expected to contain more lexical items typical of dialogue, thus chances are that the individual writers and translators will use more dialogue-prone items.

On a more general note regarding dialogue in fiction, Axelsson (2009: 191) draws attention to the fact that "[o]ne cannot take it for granted […] that the amount of direct speech or the linguistic features of direct speech are similar in samples from different parts of the fiction texts (beginning, middle and end samples)." In the present context, however, the samples are considered homogeneous and therefore comparable, as all the texts in the original ENPCfiction are taken from the beginning of texts, while the texts making up the extension are all full texts.

Axelsson (ibid.: 191-192) also mentions the lack of mark-up of direct speech in existing corpora as a problem. Although such mark-up is in fact available in the original ENPCfiction, a systematic discussion of items typically found in dialogue vs. narrative was not considered essential to the current analysis.

The results of this study seem to paint a more complex picture than the other revisited case studies regarding the impact of corpus size in contrastive studies. Although the size of the corpora definitely plays a role in that a formidable increase in the number of attested instances of *probably* and *antakelig* can be noted, the increase may also be attributed to individual preferences and genre. In other words, there is no evidence that corpus size alone accounts for the differences in distribution of these items in ENPCfiction vs. the ENPC+, suggesting that the original observations made on the basis of the smaller ENPCfiction may still be valid for the two languages in question.

4. Corpus size in contrastive studies: Summing up

In order to discuss the implications of corpus size in contrastive studies, it is important to return to the underlying questions and concerns that triggered the present study: Does corpus size matter?; and is the original ENPC large enough to yield valid results?

The experiment of replicating previous ENPC studies on the basis of a bigger corpus has thrown up various, and perhaps not new and unexpected, issues related to corpus size and the nature of corpora in general. However, corpus size alone may have less impact than was perhaps feared. Given the design and structure of the original ENPC, where care was taken to include as many different writers and translators as possible, we can rely on the findings even if the material feels limited in terms of number of tokens. Nevertheless, it should be noted that bigger parallel corpora are of course needed for studies of less frequent items and constructions, as well as for studies of recurrence. Moreover, as noted by Johansson (2011: 128):

> Although there are many advantages with the bidirectional translation model, corpora built in this way may need to be supplemented by larger corpora compiled according to the two main models presented above [comparable monolingual corpora and unidirectional translation corpora], as these are less constrained with respect to the types and range of texts.

It has also been confirmed that the issue of the individual is far from a trivial one; the need to keep an eye on dispersion and potential rogue texts is certainly upheld, particularly, it seems, in studies of lexis. The contrastive nature of the studies reported on here is of course a complicating factor, as it involves two versions of the same text, viz. the original and its translation, each potentially marked by the author's or translator's individual style. And, not unexpectedly, it seems as if discourse particles and adverbs of the kind discussed here are more easily subject to individual preferences both on the part of the writers and the translators, and are thus more likely to skew the results. In addition, lexical choices may also be influenced by genre, as in the case of items typical of dialogue in crime fiction. In this respect individual style and genre seem to be variables that have a greater impact on the results than mere corpus size.

Furthermore, we have seen that it is not unproblematic to expand parallel corpora in the way done for the ENPC+, particularly because of the time lag between the texts in the original ENPC and those making up the expansion. This was seen as a particularly decisive factor in the love-hate study in Section 3.2.

The investigation started out by asking whether corpus size matters, and the first previous ENPC study that was probed into, on the English

correspondences of *bli* + past participle (Section 3.1), showed that it does not necessarily matter. In other words, corpus size had little effect on the findings and conclusions similar to those drawn in the original study could also be drawn on the basis of the larger data set.

However, the studies revisited in Sections 3.2.and 3.3 resulted in a more complex answer to the initial question, involving, it could be claimed, different kinds of rogueness. Figure 2 is an attempt at capturing (some of) the multi-faceted nature of such rogueness.
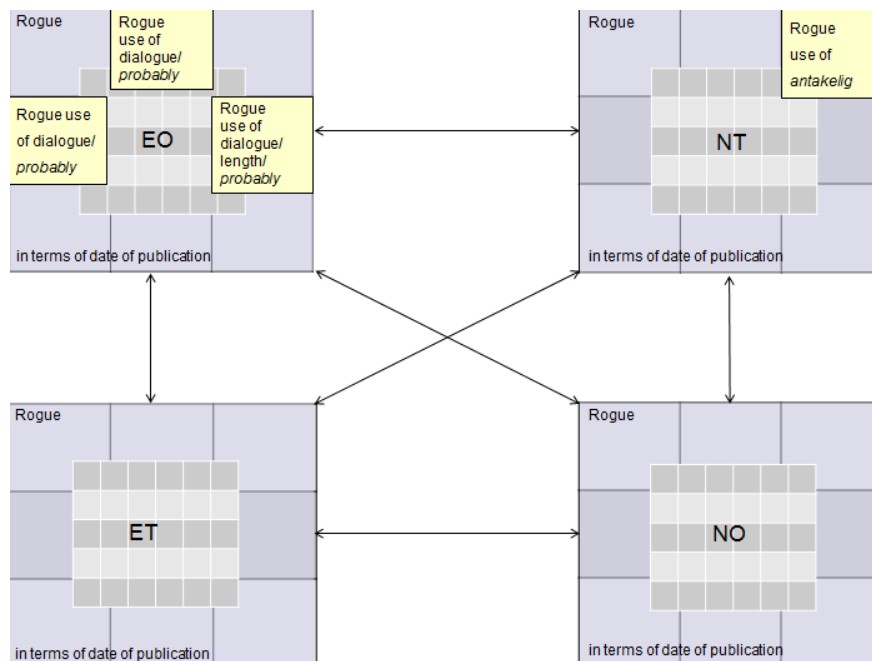

Figure 2. Some types of rogueness identified in the ENPC/ENPC+, including date of publication, use of specific lexical items, use of dialogue and length of text

In the second study (Section 3.2), the findings on the use of *love* and *elske* were found not to be affected by corpus size or any kind of rogueness, while the findings for *hate* and *hate* suggested that a language change has taken place or, indeed, is ongoing; thus, the more recent texts can be described as rogue in terms of publication date. This is indicated in Figure 2 by the area surrounding the 30x4 squares illustrating the

original ENPCfiction; all of the nine full texts constituting the extension – represented by the larger squares in Figure 2 – are rogue in terms of publication date.

The final study (Section 3.3) was the one that really drew our attention to rogueness and the role of the individual, i.e. the writer and the translator. Rogueness in the use of specific lexical items was identified, as well as rogueness in the use of dialogue in three crime fiction texts in particular; additionally one of those texts is longer than the other texts in the corpus. These issues are illustrated (in Figure 2) in the squares representing three texts in English originals (EO) and one square representing one text in Norwegian translations (NT).

With such a broad definition of rogueness as the one adopted here, a text can be rogue in one area of study (*probably*) but mainstream in another area of study (*bli* + present participle). This calls for caution on the part of (parallel) corpus compilers in the future; it will be all the more important to include a wide variety of writers and translators and to check for rogueness in each study that is carried out.

5. Conclusion and future prospects

Even if this study has highlighted some problematic issues related to corpus size and structure, these can to a large extent be controlled for, and contrastive studies based on existing parallel corpora still yield valid and sound results. However, as hinted at in Section 3.3, corpus-based contrastive studies would in the future benefit from more sophisticated statistical treatments for analysing the impact of the different variables involved. One potential variable that may play a role when it comes to linguistic choices is language variety, e.g. British vs. American English.[10] This was mentioned when outlining the differences between ENPCfiction and its extension but was not discussed in connection with the studies revisited. It is, nevertheless, a factor worth taking into consideration, particularly if robust statistical tests were implemented.

Moreover, Axelsson (2011), in her cross-linguistic study of tag questions, points to another restriction on the data culled, in her case, from the ESPC, namely the fact that only samples from the beginning of

---

[10] See also Axelsson (2011: 219), who draws attention to this as a factor potentially contributing to certain preferred usage patterns.

texts are included. Thus, following Sinclair (2005: 6), it could be argued that "ideally, documents and transcripts of verbal encounters should be included in their entirety". However, copyright holders rarely donate complete texts to a corpus, thus opening up for including different parts of texts, e.g. beginning, middle and concluding samples. This would, unfortunately, introduce yet another variable that may or may not prove significant in the study of language use.

With a pragmatic approach to what is feasible to obtain, Johansson seems to have struck the right balance for devising an appropriate structure for bidirectional parallel corpora (see Johansson & Hofland 1994; Johansson et al. 1999/2002). As a future prospect, it should therefore be encouraged to compile parallel corpora matching the existing ones in terms of content and structure, but comprising texts of a more recent date. In a similar fashion to what has been done for the LOB and Brown corpora – with FLOB and Frown[11] – a carefully designed ENPC 20 years on would pave the way for a new field of diachronic corpus-based contrastive studies, ensuring that such studies can be carried out in a systematic way.

References
Altenberg, Bengt. 1999. "Adverbial connectors in English and Swedish: Semantic and lexical correspondences." *Out of Corpora: Studies in Honour of Stig Johansson*. Eds. Hilde Hasselgård & Signe Oksefjell. Amsterdam: Rodopi. 249–268.
Axelsson, Karin. 2009. "Research on fiction dialogue: Problems and possible solutions." *Corpora, Pragmatics and Discourse*. Eds. Andreas H. Jucker, Daniel Schreier & Marianne Hundt. Amsterdam: Rodopi. 189–201.
Axelsson, Karin. 2011. *Tag Questions in Fiction Dialogue*. Ph.D. thesis. Göteborg: University of Gothenburg. [Available online at http://hdl.handle.net/2077/24047]

---

[11] The Lancaster-Oslo/Bergen (LOB) and Brown corpora are carefully designed 1-million-word monolingual corpora of British and American English, respectively, containing texts from 1961. Their 1991 counterparts – FLOB and Frown – were compiled according to the same criteria at Freiburg University, hence the 'F', to enable diachronic studies of British and American English (see the ICAME corpus collection at http://clu.uni.no/icame/newcd.htm).

de Haan, Pieter. 1992. "The optimum corpus sample size?" *New Directions in English Language Corpora*. Ed. Gerhard Leitner. Berlin / New York: Mouton de Gruyter. 3–19.

de Haan, Pieter. 1996. "More on the language of dialogue in fiction." *ICAME Journal* No. 20, 23–40.

Ebeling, Jarle & Signe Oksefjell Ebeling. 2013. *Patterns in Contrast*. Amsterdam: John Benjamins.

Ebeling, Jarle, Signe Oksefjell Ebeling, & Hilde Hasselgård. 2013. "Using recurrent word-combinations to explore cross-linguistic differences." *Advances in Corpus-based Contrastive Linguistics. Studies in Honour of Stig Johansson*. Eds. Karin Aijmer & Bengt Altenberg. Amsterdam: John Benjamins. 177–200.

Ebeling, Signe Oksefjell. 2003. *The Norwegian Verbs bli and få and their Correspondences in English. A Corpus-based Contrastive Study*. Oslo: Acta Humaniora.

Ebeling, Signe Oksefjell. 2015a. "Loving and hating in English and Portuguese: A corpus-based contrastive study." *Linguística, Informática e Tradução: Mundos que se Cruzam. Homenagem a Belinda Maia*. Eds. Alberto Simões, Anabela Barreiro, Diana Santos, Rui Sousa-Silva & Stella E. O. Tagnin. *Oslo Studies in Language* (OSLa) 7(1): 439–456.

Ebeling, Signe Oksefjell. 2015b. "A contrastive study of Norwegian pseudo-coordination and two English posture-verb constructions". *Cross-linguistic Perspectives on Verb Constructions*. Eds. Signe Oksefjell Ebeling & Hilde Hasselgård. Newcastle upon Tyne: Cambridge Scholars. 29–57.

Granger, Sylviane. 1998. "The computer learner corpus: a versatile new source of data for SLA research." *Learner English on Computer*. Ed. Sylviane Granger. London / New York: Longman. 3–18.

Hasselgård, Hilde. 2011. "Loving and hating in English and Norwegian speech." Paper presented at the Jan Svartvik Birthday Symposium, Lund, 19 August 2011.

Johansson, Mats. 1996. "Contrastive data as a resource in the study of English clefts." *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies, Lund 4-5 March 1994*. Eds. Karin Aijmer, Bengt Altenberg & Mats Johansson. Lund: Lund University Press. 127–152.

Johansson, Stig. 1998a. "On the role of corpora in cross-linguistic research." *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Eds. Stig Johansson & Signe Oksefjell. Amsterdam: Rodopi. 3–24.

Johansson, Stig. 1998b. "Loving and hating in English and Norwegian: A corpus-based contrastive study." *Perspectives on Foreign and Second Language Pedagogy*. Eds. Dorte Albrechtsen, Birgit Henriksen, Inger M. Mees & Erik Poulsen. Odense: Odense University Press. 93–103.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies.* Amsterdam/Philadelphia: John Benjamins.

Johansson, Stig. 2008. *Contrastive Analysis and Learner Language: A Corpus-based Approach.* Compendium for the course ENG2152 – Contrastive and Learner Language Analysis, University of Oslo.

Johansson, Stig. 2009. "Which way? On English *way* and its translations." *International Journal of Translation* 21(1–2): 15–40.

Johansson, Stig. 2011. "A multilingual outlook of corpora studies." *Perspectives on Corpus Linguistics*. Eds. Vander Viana, Sonia Zyngier & Geoff Barnbrook. Amsterdam/Philadelphia: John Benjamins Publishing Company. 115-129.

Johansson, Stig, Jarle Ebeling & Signe Oksefjell. 1999/2002. "English-Norwegian Parallel Corpus: Manual." www.hf.uio.no/ilos/english/ services/omc/enpc/ENPCmanual.pdf

Johansson, Stig & Knut Hofland. 1994. "Towards an English-Norwegian Parallel Corpus". *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora, Zurich 1993*. Eds. Udo Fries, Gunnel Tottie, & Peter Schneider. Amsterdam: Rodopi. 25–37.

Johansson, Stig & Berit Løken. 1997. "Some Norwegian discourse particles and their English correspondences." *Sounds, Structures and Senses. Essays Presented to Niels Davidsen-Nielsen on the Occasion of his Sixtieth Birthday*. Eds. Carl Bache & Alex Klinge. Odense: Odense University Press. 149–170.

Oksefjell, Signe. 1999. "A description of the English-Norwegian Parallel Corpus: Compilation and further developments." *International Journal of Corpus Linguistics* 4:2. 197–219.

Sinclair, John. 2005. "Corpus and text – Basic principles." *Developing Linguistic Corpora: A Guide to Good Practice*. Ed. Martin Wynne. Oxford: Oxbow Books. 1–16. [Available online at http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm]

Viberg, Åke. 2010. "Basic verbs of possession: A contrastive and typological study." *ConTextes* Vol. 4 – *Unison in Multiplicity: Cognitive and Typological Perspectives on Grammar and Lexis*. http://cognitextes.revues.org/308

The way of the world: the colligational framework "the N1 of the N2" and its Norwegian correspondences

Hilde Hasselgård, University of Oslo

Abstract
The English-Norwegian Parallel Corpus is used to study the colligational framework "the N1 of the N2" (e.g. *the way of the world*). The aim is to discover recurrent patterns in the lexical and semantic make-up of such sequences in English and their correspondences in Norwegian. The correspondences are described in terms of both structure and meaning. In spite of similarities in the structural potential of English and Norwegian noun phrases, the majority "the N1 of the N2" sequences have divergent correspondences in Norwegian. This points to differences between the languages in their preferred lexicogrammatical patterns. The study confirms the feasibility of carrying out cross-linguistic studies on the basis of patterns of function words.

1. Introduction

This paper explores the colligational framework "the N1 of the N2" and its Norwegian correspondences.[1] In so doing, it represents an attempt at extending the domain of contrastive studies to patterns based on function words. The term "colligational framework" is modelled on Renouf & Sinclair's "collocational framework", defined as a discontinuous sequence of (grammatical) words (Renouf & Sinclair 1991: 128). Since it is not grammatically complete (or "self-standing"), its "wellformedness is dependent on what intervenes" (ibid.). Unlike Renouf & Sinclair's study, the present one specifies word class membership for the two empty slots, hence the modified term. Renouf & Sinclair demonstrate that such frameworks are highly selective of their collocates. For example, focusing on the intermediate word in the framework "a _ of", they find that the quantifiying nouns *couple, series, pair* and *lot* occur in this framework more frequently than they occur outside it (ibid.: 131).

Hunston (2008) argues that "small words" play an important role in the identification of grammar patterns, which in turn can form semantic sequences, i.e. "series of meaning elements that can be demonstrated to

---

[1] I am grateful to the reviewers of this paper for constructive criticism, and in particular for suggesting the term "colligational framework".

occur regularly" (2008: 271). That is, the core of semantic sequences need not be lexical words, since grammatical words have also been shown to have collocates; see e.g. Groom (2007), who studies the phraseological profiles of a number of grammatical words. Groom finds, for instance, that in a random sample of 100 instances of *of*, a number of semantic sequences can be recognized, such as PROPERTY + *of* + PHENOMENON, illustrated by *the essential values of academic life* (2007: 84). As also noted by Hunston, prepositions can "serve to classify semantically the lexical words with which they occur" (2008: 202). Such semantic classifications are of great interest in a contrastive study.

In the British National Corpus (BNC) *the end of the* and *the rest of the* are both among the ten most frequent 4-grams.[2] This suggests that "the N1 of the N2" is a common and productive colligational framework worth investigating contrastively. In addition, complex noun phrases are intriguing from an English-Norwegian contrastive perspective: as Elsness (2014: 91) observes, "there are some notable differences in the structure of the noun phrase between the two languages", particularly in the use of modifiers. In spite of this, contrastive studies of complex noun phrases in English and Norwegian are scant, Elsness (2014), focusing on clausal modifiers, being an exception.

This study explores the colligational framework "the N1 of the N2" with the aim of discovering recurrent patterns as regards the selection of lexical items in the frame as well as its Norwegian correspondences. The investigation of English will focus on the patterns that emerge with regard to the types of nouns in the open slots and the relation between the N1 and N2. The Norwegian correspondences of the pattern will be studied in order to discover the extent to which Norwegian uses similar or different patterns, and whether different semantic types of nouns, or combinations of nouns in the "the N1 of the N2" framework, trigger different structural types in Norwegian. Because of the exploratory nature of the study, and in order to limit the material, the investigation is limited to definite noun phrases, thus ignoring e.g. the related framework "a(n) _ of" studied by Renouf & Sinclair (1991). It is hoped not only that this limited investigation will point to systematic similarities and differences between English and Norwegian complex noun phrases, but

---

[2] According to a search using "Phrases in English": http://phrasesinenglish.org, accessed 29 July 2015.

also that such an investigation will demonstrate the usefulness of starting an investigation from grammatical words and exploring colligational frameworks in contrast.[3]

2. Complex noun phrases in English and Norwegian

English and Norwegian noun phrases have much the same syntactic structure, with one major exception: the definite form of Norwegian nouns is marked morphologically by a suffix. The definite suffix usually replaces the definite article, but if the noun is premodified, the noun phrase typically (but not obligatorily) has double definiteness, with both article and definite suffix. Examples are shown in (1), in which the definite suffixes have been underlined.[4]

(1)　On *the way* to *the hotel* where he was staying, *the emaciated man*
　　　walked along the harbour of *the little town*. (FC1T)
　　　På *veien* til *hotellet*, der han bodde, gikk *den radmagre mannen*
　　　langs *havnen* i *den lille byen*. (FC1)
　　　Lit: "On way$_{def}$ to hotel$_{def}$, where he lived, walked the emaciated
　　　man$_{def}$ along harbour$_{def}$ in the little town$_{def}$.

As the example also shows, nouns in both languages can be premodified by adjectives and postmodified by prepositional phrases. Nominal premodification is not common in Norwegian unless the premodifying noun has genitive form; noun + noun combinations in English may correspond to compounds in Norwegian; see (2).

(2)　It was *Harry's idea* that they should give a *dinner party* … (FW1)
　　　Det var *Harrys idé* at de skulle ha *middagsselskap* … (FW1T)
　　　Lit: "It was Harry's idea that they should have dinnerparty"

---

[3] Miccoli (2010) studied phrasal verbs in English and Norwegian on the basis of lexically specified particles; this is the only previous study to my knowledge that explores collocations on the basis of function words. Brems (2015) studied quantifying binomials with *of*, but with a lexically specified N1.

[4] All examples are from the English-Norwegian Parallel Corpus. The English version consistently appears first. If it is a translation, its identification tag ends in 'T'. The Norwegian version is immediately followed by a literal translation marked as "Lit.". The most relevant parts of the examples have been italicised.

It may be noted that genitive expressions are also similar across the languages: like English, Norwegian has an *s*-genitive, illustrated in (2), as well as an equivalent of the *of*-genitive, using the preposition *til* (e.g. *the name of the journalist – navnet til journalisten*). For further details of complex noun phrase structure in English and Norwegian, see e.g. Biber et al (1999: 574 ff) and Faarlund et al. (1997: 233 ff). The structural similarities are great enough for the "the N of the N" construction to have congruent correspondences in Norwegian, allowing for the replacement of the definite article by the definite suffix, as seen in (3).

(3)   I live in that caravan on *the edge of the cliff*. (PDJ3)
        Jeg bor i campingvognen der borte på *kanten av skrenten*. (PDJ3T)
        Lit: "I live in caravan$_{def}$ there over on edge$_{def}$ of cliff$_{def}$."

However, the similarities in syntactic structure, and thereby in constructional potential, are "no guarantee that there is identity of use" (Johansson 2012: 47). Rather, as is often found in contrastive studies, it is likely that the two languages differ in their "preferred ways of expressing similar meanings" (ibid.: 64).

3. The colligational framework "the N1 of the N2"
Sequences that instantiate the framework "the N1 of the N2" are complex noun phrases, according to most accounts consisting of a head noun (N1) plus a postmodifier (*of* N2). According to Biber et al. (1999: 606), prepositional phrases are the most frequent type of postmodifier of nouns, and of these, 60-65% are introduced by *of* (ibid.: 635). Furthermore, lexical bundles involving *of*-phrases are highly frequent in academic prose: a list of four-word bundles of this type contains numerous instances such as *the end of the, the beginning of the, the base of the* (ibid.: 1014 f). Some such bundles are also noted for conversation, particularly in temporal and spatial references such as *the end/middle of the week/road* (ibid.: 1012).

    However, *of* is not a "typical" preposition; in particular it generally does not introduce PPs with an adjunct function (Sinclair 1991; Owen 2007). According to Sinclair, "it may ultimately be considered distracting to regard *of* as a preposition at all" (1991: 83). On the other hand, *of* is highly multifunctional in terms of the meaning relationships it

expresses between nouns preceding and following it. In addition, an "N *of* N" sequence is not necessarily one of head noun followed by a postmodifier; according to Sinclair, the second noun seems to be most salient. Hence, *of* can introduce a second noun as a potential headword in phrases such as *this kind of problem, the bottle of port* (1991: 85), an issue also discussed by Keizer (2007).

The meaning relations expressed by *of* in the "N of N" sequence can to some extent be linked to the meaning of the N1 according to Sinclair (1991: 87 ff). The N1 can be a "focus noun", specifying some part of the N2 (which is regarded as the head noun), e.g. *the top of the pillar*. The focus can also be on a "specialized part", e.g. *the first week of the war, the horns of the bull*, or on "a component, aspect, or attribute", e.g. *a gasp of shock, the study of geography*. Alternatively, the N1 can be a "support noun", often reduced in meaning and "offering some kind of support to N2" (ibid: 89), e.g. *the notion of machine intelligence, an object of embarrassment*.

Keizer (2007), whose primary interest is in the structural properties of the noun phrase, takes a different perspective on the semantic relations expressed by *of*. "Possession" is seen as a cover term that includes ownership (of material belongings as well as more abstract products), kinship and body parts. Possession is also extended to non-human possessors, e.g. *the mosaics of Venice, the problems of the world* (2007: 63). A particular group of nouns in the N1 position are called "relational" (ibid.: 64): relational nouns include kinship terms (*the son of Aron*), body parts, and "nouns which denote parts of (physical or abstract) features of entities", such as *size, middle, feature, nature* (ibid.); thus nouns that might be seen as either focus nouns or support nouns in Sinclair's (1991) framework. Other functions of *of*-phrases recognized by Keizer (2007: 65, 73) are partitive (*the first of a few questions*) and appositional (*the job of foreign minister*); see also Quirk et al. (1985: 1284).

For the present analysis the semantic relations were classified descriptively in terms of an analysis of the N1 into the following categories (which might also be seen as semantic sequences; cf. Hunston 2007; Groom 2007):

- part of N2, e.g. *the corners of the tea-towel*
- locative, e.g. *the back of the bus*
- temporal, e.g. *the night of the ball*

- owned/caused by N2, e.g. *the noise of the TV*
- body part, e.g. *the legs of the men*
- feature of N2, e.g. *the names of the places, the colours of the rainbow*
- person, e.g. *the leader of the expedition*
- nominalization, e.g. *the outbreak of the Plague*
- support, e.g. *the idea of the echo*
- ordinal, e.g. *the last of the fishfingers*
- N1 contains N2, e.g. *the pictures of the frog*

Admittedly, there is some overlap between the categories. In particular, it can be argued that the categories "locative", "temporal" and "body part" are superfluous because they are subtypes of the "N1 is part of N2" relation. However, for the purposes of this study, also taking the Norwegian correspondences into account, it makes sense to single out these subtypes. Several of the categories might be grouped under "possession" in Keizer's (2007) wide sense, i.e. the top six on the list, and possibly also nominalization (2007: 62). The category of "support noun" has been borrowed from Sinclair (1991). His category of "focus noun", however, has not been applied, as a more fine-grained system was needed for this study, e.g. one that can single out locative nouns. Support nouns would correspond to Keizer's appositional relation, and ordinals to the partitive one. The list presented above seems to make sense in light of the correspondences that occur; see further section 5.

4. Material and method

The investigation is based on the fiction part of the English-Norwegian Parallel Corpus (ENPC). The ENPC is a bidirectional translation corpus containing original texts in both English and Norwegian with translations into the other language. See Johansson et al. (1999/2002) for further details of the corpus.

For the present study, the raw text files of English originals and translations were imported into WordSmith Tools, and the string *the * of the* * was searched for. My interest was in recurrent patterns, but since very few sequences recurred in identical form, sorting and selection were made according to the N1 in the sequence following the observation that there was more recurrence of N1s than of N2s. All single occurrences

were removed, as were those occurring in one text only, leaving 430 hits from English originals and 507 from translations. These were matched up with their Norwegian correspondences, stored in a database and annotated for the lexemes occurring as N1 and N2 and the semantic relation between them, correspondence type (congruent, divergent or zero; cf. Johansson 2007: 25), and structural type of the Norwegian correspondence.[5]

In this context correspondences were considered congruent (i.e. structurally equivalent) if they consisted of two definite nouns and an intervening preposition, as in (4). Divergent correspondences are all other structural types, e.g. compounds, as in (2) above, or word class change for either the N1 or the N2, as shown in (5), where the N1 plus the preceding preposition correspond to an adverb.

(4)   No identification on *the back of the photograph*. (NG1)
       Men det finnes ingen identifikasjon på *baksiden av fotografiet*.
       (NG1T)
       Lit: "But there exists no identification on *backside*$_{def}$ *of*
       photograph$_{def}$."

(5)   A yellowed newspaper cutting lay *on the top of the pile*. (GS1T)
       Øverst i bunken lå et gulnet avisutklipp. (GS1)
       Lit: "*Uppermost in pile*$_{def}$ lay a yellowed newspapercutting."

There are also some cases of zero correspondence, which occurs if either the "N1 of the N2" sequence or the relevant s-unit has no counterpart in the Norwegian text. Cases where either the N1 or the N2 is missing from the Norwegian source or orignal, however, are analysed as divergent correspondences (of the structural types "N1/N2 missing").

5. Corpus investigation
This section presents an analysis of the colligational framework "the N1 of the N2" and its Norwegian correspondences. Section 5.1 surveys the noun collocates of the framework before the correspondences are

---

[5] Note that the term "correspondence" covers both sources and translations, c.f. Johansson (2007: 23).

explored with regard to congruence and divergence, structural types of Norwegian correspondences, meanings of the N1s, and individual patterns for the most frequent N1s. It is expected that recurrent nouns in the framework will form certain (types of) semantic sequences. Assuming that Norwegian is less fond of complex noun phrases than English is (cf. Elsness 2014), congruent correspondences will be more frequent in Norwegian translations than in Norwegian originals. Furthermore, if Sinclair is right about the greater prominence of the N2 in such sequences, N1 can be expected to be divergent more often than N2. It is likely that individual N1 lexemes or combinations of N1 and N2 as well as certain meaning relations may come with their preferred correspondence types. Preferred correspondence types will thus be related to both patterns and individual nouns.

5.1 Nouns occurring in the "the N1 of the N2" framework

The nouns occurring above 10 times in the N1 position are shown in Table 1. Singular and plural forms of the same noun have been grouped under the same lemma, as have the British and American spellings of *centre/center*.

There is a reasonable degree of overlap between the N1s occurring in this pattern in originals and translations, and those nouns that are shown to occur only in originals in Table 1 are to be found among the top 20 of the translations and vice versa. The majority of the words in both columns are locative and/or denote a part of whatever the N2 refers to. As will be detailed in Section 5.3, "part of N2" and "locative" are indeed by far the most frequent categories of N1.

Table 1. N1 collocates selected by the framework (min. 10 occurrences)

	English original		**English translation**	
rank	lemma	frequency	lemma	frequency
1	end	39	middle	52
2	rest	26	edge	43
3	back	22	rest	30
4	edge	22	end	23
5	side	18	back	16
6	centre	14	bottom	15
7	top	14	side	14
8	middle	13	top	14
9	bottom	12	sound	13

Only a handful of N2s occur above 10 times, and the majority of those refer to places and times, as shown in Table 2. The most frequent N2s are, in short, nouns that go together with the most frequent N1s, i.e. they denote things that have middles, edges, sides or ends. The impression from Tables 1 and 2 combined is that the framework "the N1 of the N2" is associated with locative, and to some extent temporal, expressions (cf. Biber et al. 1999: 1012).

Table 2. N2 collocates selected by the framework (min. 10 occurrences)

	English original		**English translation**	
rank	lemma	frequency	lemma	frequency
1	house	15	table	15
2	road	11	room	14
3	room	10	house	11
4	(table	8)	night	11
5	(car	7)	day	10

5.2 Congruence and divergence

The overall frequencies of the correspondence types displayed in Table 3 indicate that the first hypothesis was correct: the proportion of congruent correspondences is higher when English is the source language. The difference between translations and sources is significant (Fisher's exact test, P value = 0.0009).

The fact that divergent correspondences are more frequent than congruent ones in both directions shows that the construction represents a point where the languages differ in spite of the similarities in syntactic repertoire pointed out in Section 2.

Table 3. Congruence and divergence

	E orig → N tran		E tran ← N orig)	
	N	%	N	%
congruent	187	43.5	167	32.9
divergent	232	54.0	331	65.3
zero	11	2.6	9	1.8
total	430	100	507	100

The higher frequency of congruent correspondences in Norwegian translations may be due to the possibility of using the "N prep N" construction in Norwegian in expressions where other structures may be more common. Example (6) may be an illustration of this; although the translation is correct, the compound *leirlederen* is a more idiomatic alternative.[6]

(6)   *The director of the camp*, not wanting to break the news on the
       phone, had driven to Baltimore to tell them in person. (AT1)
       Lederen av leiren, som ikke hadde lyst til å fortelle nyheten i
       telefonen, hadde kjørt til Baltimore for å fortelle det personlig.
       (AT1T)
       Lit: "Leader$_{def}$ of camp$_{def}$ …"

It may be noted that "the N1 of the N2" occurs as the complement of a preposition in about 65% of the cases in both originals and translations. Norwegian translations from English have similar proportions of congruent and non-congruent correspondences both within and outside prepositional phrases (just under half are congruent, and just over half divergent). In contrast, Norwegian sources tend to diverge from English translations if the pattern is part of a PP: 25% congruent and 73%

---

[6] In the 70-million-word *Leksikografisk bokmålskorpus* ("Lexicographical corpus of Norwegian"), *leirleder* occurs 8 times and *leder(en) av leiren* not at all. A Google search found both forms, but *leirlederen* was more than ten times as frequent.

divergent, as compared to equal shares of congruent and divergent correspondences if the sequence is not preceded by a preposition.

Figure 1 shows the types of divergences that occur and the percentage of each (from among the total number of divergent correspondences). The N1 is the most frequent locus of divergence: in Norwegian sources (of English translations), the N1 is divergent five times as often as the N2, and in Norwegian translations (of English originals), the N1 is divergent eight times as often as the N2. It is interesting to note that particularly compounds and adverbs corresponding to the English N1 are more frequent in Norwegian originals while the Norwegian translations have larger shares of the N1 missing and of clausal expansions of the "the N1 of the N2" construction.
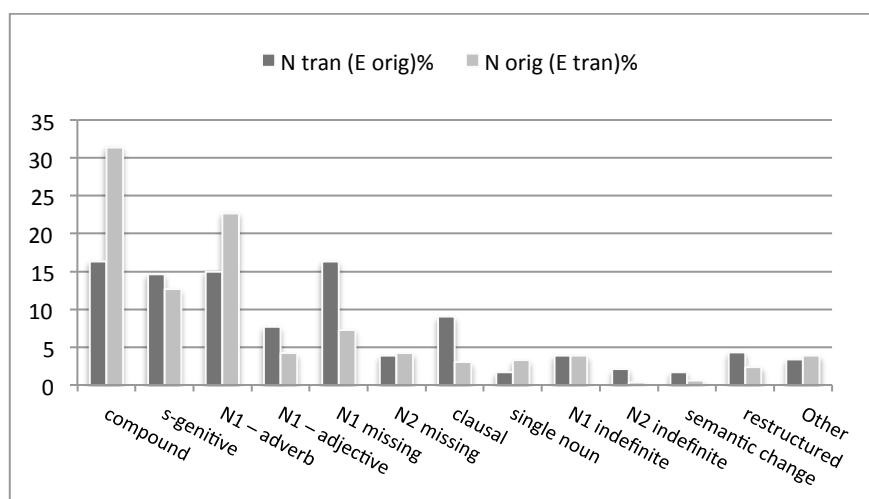

Figure 1: Structural types of divergent correspondences of the "the N1 of the N2" (percentages)

Examples (7)-(13) show the most common types of divergence. In (7) the compound noun *hushjørnet* has been translated by "the N1 of the N2", while in (8) the English *of*-genitive is rendered by a Norwegian *s*-genitive. Both examples illustrate potentially systematic differences between the two languages that deserve further study: my hypothesis at this point would be that both compounding and the *s*-genitive are more widespread and productive in Norwegian than they are in English.

(7)   She disappeared round *the corner of the house* … (THA1T)
        Hun forsvant rundt *hushjørnet* … (THA1)
        Lit: "She disappeared round housecorner$_{def}$"

(8)   I'm *the head of the household*. (ST1)
        Jeg er *familiens overhode*. (ST1T)
        Lit: "I am family$_{def}$'s head."

In examples (9) and (10) the English N1 corresponds to an adverb or an adjective in the parallel Norwegian phrase. It should be noted that the phrase *in the middle of* corresponds quite regularly (though not invariably) to the Norwegian adverb *midt* followed by the preposition *i* ('in') or *på* ('on'), and it may be argued that both *in the middle of* and *midt i/på* are complex prepositions. The same pattern is, however, found with other locative nouns, as illustrated by (5) above, where (*on*) *the top of* corresponds to *øverst* ('uppermost'). In the correspondences where the English N1 has been rendered by a Norwegian adjective, shown in (10), it is hard to see any clear patterns.

(9)   What was I doing here in *the middle of the afternoon*? (KF1T)
        Hva gjorde jeg her *midt på ettermiddagen*? (KF1)
        Lit: "What did I here in-the-middle on afternoon$_{def}$"

(10)   It was *the end of the afternoon*… (JB1)
        Det var *sen ettermiddag*… (JB1T)
        Lit: "It was late afternoon…"

In examples (11) and (12) either the N1 or the N2 is missing from the Norwegian version. The N1s most frequently missing from the Norwegian are locative, as in (11) – the same phenomenon is seen for e.g. *the bottom of, the side of*. The missing N2s, on the other hand, tend rather to be either retrievable or inferable from the context, as in (12), where the English translator has deemed it necessary to add the specification *of the street*. This is not, however, simply caused by the translation feature of explicitation, as the reverse phenomenon is equally common in translation from English into Norwegian.

(11) We stared at one another across *the expanse of the room*. (TH1)
      Vi stirret på hverandre gjennom *rommet*. (TH1T)
      Lit: "We stared at eachother through room$_{def}$."

(12) Brita and Hildegun stopped at *the corner of the street*, dropped
      their schoolbags onto the pavement and sat on the wall. (BV1T)
      Brita og Hildegun stanset på *hjørnet*, slapp skoleveskene på
      fortauet og satte seg på muren. (BV1)
      Lit: "Brita and Hildegun stopped at corner$_{def}$…"

All in all, the high proportion of divergence and the types of changes that
occur in translation between the languages seem to indicate that
Norwegian is less favourable than English is to complex noun phrases of
the type "definite noun + preposition + definite noun".

5.3 Meanings of "the N1 of the N2" and their correspondences
This section looks at the broad meaning categories of N1s outlined in
Section 3 to see if they vary in their preferred correspondence types. The
main results are shown in Table 4, in which zero correspondences have
been ignored due to their low numbers (cf. Table 3).

Table 4. Variation across broad N1 meanings (raw frequencies)

	English original	**English translation**
	Congruent / divergent	Congruent / divergent
part of N2	88 / 56	58 / 108
locative	38 / 83	21 / 126
owned/caused by N2	18 / 29	36 / 32
person	8 / 13	11 / 10
feature	6 / 13	5 / 11
body part	5 / 10	3 / 5
nominalization	6 / 14	11 / 20
support	7 / 8	8 / 13
temporal	6 / 4	8 / 2
N1 contains N2	-	6 / 2
ordinal	5 / 2	0 / 3

Most of the meaning categories have a majority of divergent
correspondences with the important exception of "part of N2" in English

originals. In English translations, however, divergent correspondences are almost twice as frequent as congruent ones within this meaning category. This suggests that English and Norwegian have different preferences in expressions of part-whole relationships even though the same structures may be available. Examples (13) and (14) illustrate a congruent and a divergent correspondence of "N1 is part of N2", respectively.

(13)   And then *the door of the studio* is opened and several people
         enter… (ABR1)
         Og så blir *døren til atelieret* åpnet, og flere mennesker kommer
         inn… (ABR1T)
         Lit: "And then is door$_{def}$ to studio$_{def}$ opened …"

(14)   The planet has already gone a good way across *the surface of the
         sun* and will soon pass the center. (EFH1T)
         Planeten er kommet et godt stykke lenger inn over *soloverflaten*
         allerede, snart passerer den midten. (EFH1)
         Lit: "Planet$_{def}$ is come a good bit longer in over sun-surface$_{def}$
         already…"

The Norwegian original in (14) uses a compound noun instead of the complex noun phrase framework, which is the most frequent structural type of divergence in the category of part-whole relationships. Simplification of the noun phrase in terms of omission of either N1 or N2 or the use of a noun simplex also occurs, e.g. *the foot of the garden – hagen* ('the garden'); *the corner of the room – kroken* ('the corner'); *details of the case – opplysninger* ('information').

Sequences with a locative N1 form the meaning category with the most consistently divergent correspondences. The most common type of divergence is correspondence between an English N1 and a Norwegian adverb. This happens in 59% (74 out of 126) of the divergent locative expressions in Norwegian sources of English translations, and in 39% (32 out of 83) of Norwegian translations. Examples are given in (5) and (9) above. The "the N1 of the N2" framework with locative N1 also frequently corresponds to N1 omission in Norwegian translations (21 out of 83) and to compounds (11 out of 83 in translations and 20 out of 126 in originals). Other divergence types are less frequent.

N1s denoting something that is owned or caused by the N2 have a relatively large proportion of congruent correspondences, especially in Norwegian originals where they outnumber the divergent ones. (15) and (16) show a congruent and a divergent example, respectively. The *s*-genitive, as exemplified in (16), is the most common divergent type within this meaning category in both originals and translation, which is not surprising given the general meaning of possession.

(15)   … and she was pulled from the road into *the shadow of the bushes*. (PDJ3)
       … og hun ble trukket fra veien inn i *skyggen av buskene*. (PDJ3T)
       Lit: "and she was pulled from road$_{def}$ in to shadow$_{def}$ of bushes$_{def}$."

(16)   His elder son, Robert Falcon Scott, became one of the greatest explorers in *the history of the Empire*. (KH1T)
       Den eldste sønnen, Robert Falcon Scott, ble en av de store oppdagere i *Imperiets historie*. (KH1)
       Lit: "…one of the greatest explorers in Empire$_{def}$'s history."

N1s denoting a person have almost equal numbers of congruent and divergent correspondences in both directions of translation. No particular type of divergent correspondence dominates, but compounds and *s*-genitives occur in both originals and translations. *S*-genitives are also a common divergent correspondence of N1s denoting body parts, especially in Norwegian translations from English. When the N1 is a nominalization, clausal expansion is the most common type of divergence in Norwegian translations from English, but in the other direction of translation, compounds are most common, as in (17).

(17)   *The drone of the motor* is powerful and pleasant, ... (EH1T)
       Motorduren er sterk og god, ... (EH1)
       Lit: "Motordrone$_{def}$ is strong and good…"

In the remaining meaning categories there are no discernible patterns as regards preferred types of divergent correspondences.

5.4 Structures and meanings in some correspondence types
This subsection will take a closer look at the four divergent correspondence types that differ the most between originals and translations (cf. Fig. 1), namely

- compound (more divergence in Norwegian originals);
- N1 – adverb (more divergence in Norwegian originals);
- N1 missing (more divergence in Norwegian translations);
- clausal correspondence (more divergence in Norwegian translations).

The exploration will include a scrutiny of the N1s and the meaning relations expressed in such correspondences to find out whether particular nouns or meaning types favour particular Norwegian correspondences.

A striking number of Norwegian compounds correspond to "the edge of the _": 30 out of 104 in English translations and 7 out of 38 in English originals. An example is given in (18). The source of "the edge of the _" is typically a compound ending in *-kanten* ('the edge').

(18)   He leaned forward over *the edge of the table*. (GS1T)
        Han bøyde seg fram over *bordkanten*. (GS1)
        Lit: "He leaned himself forward over table-edge$_{def}$."

Other reasonably frequent N1s with compound correspondences are *side, mouth, surface, wall* and *turn*. *Turn* occurs exclusively in *the turn of the century*, corresponding to *århundreskiftet* ('the year-hundred-shift'); thus indicating that both "the N1 of the N2" sequence and the Norwegian compound are lexicalized.

The most frequent meaning relation between N1 and N2 in the phrases corresponding to compounds is that of part-whole, i.e. the N1 denotes a part of the N2. There are also some locative N1s, but it may be argued that these nouns also denote a part-whole relationship, e.g. *the side of the ship*. We may note that the components of the Norwegian compounds typically have the reverse order of the nouns in the "N1 of

N2" framework: in a part-whole relationship, the compounds start with the whole (e.g. 'table-edge').[7]

The N1s corresponding to a Norwegian adverb are typically locative, focusing on some part of the N2, or giving a more precise location within the N2, as shown in (19) and (20). Temporal N1s may have the same type of correspondence, cf. example (9). As noted above, both locative and temporal expressions with *in the middle of* regularly correspond to the Norwegian *midt i/på,* and these expressions account for many of the N1–adverb correspondences (48 out of 75 in English translations and 9 out of 35 in English originals). Other frequent N1s are *bottom, centre* and *front*. The locative noun corresponding to an adverb typically occurs after a preposition, as in (19); thus the adverb corresponds to the preposition plus the locative noun, as shown also by (20).

(19)　A small lamp stood in *the centre of the table*, ... (EG2T)
　　　En liten lampe med gul silkeskjerm var plassert *midt på bordet*…
　　　(EG2)
　　　Lit: "A small lamp with yellow silkshade was placed in-the-middle on table$_{def}$"

(20)　*At the bottom of the hill* is the Lady with the Fleas. (LSC1T)
　　　Nederst i bakken kommer Damen med maurene, … (LSC1)
　　　Lit: "Nethermost in hill$_{def}$ comes Lady$_{def}$ with ants$_{def}$…"

Correspondences with a missing N1 equivalent are most frequent in Norwegian translations from English. No N1 lemma is omitted or added noticeably more often than others in this group. However, many such N1s are locative, of the type that Sinclair (1991: 87) calls "focus nouns", which "[specify] some part of the N2". Sinclair regards the N2 as the headword in such constructions, and such analysis may explain the omissibility of the N1; they are semantically less important than the N2 in the framework "the N1 of the N2". An example is given in (21); see also (11) above.

---

[7] The reversed order of N1 and N2 also occurs with *s*-genitive correspondences, e.g. t*he back of the letter – brevets bakside* ('the letter's backside'); *the colours of the rainbow – regnbuens farger* ('the rainbow's colours').

(21)   He shrugged and put it at *the back of the safe*. (FF1)
       Han trakk på skuldrene og la kopien inn i *safen*. (FF1T)
       Lit: "He shrugged on shoulders$_{def}$ and put copy$_{def}$ in to safe$_{def}$."

Correspondence by means of a clausal structure is shown in (22). This type of correspondence is more common in Norwegian translations than in sources, although it occurs in both types of text, cf. Figure 1.

(22)   Mattie looked for *the direction of the voice*. (GN1)
       Mattie snudde seg for å se *hvor stemmen kom fra*. (GN1T)
       Lit: "Mattie turned herself for to see where voice$_{def}$ came from."

N1s in frameworks with clausal correspondences include those that can be read as nominalizations, as in *the outbreak of the punk vogue – da punk-bølgen brøt ut* ('when the punk-wave broke out'), and those that reflect subjective genitives (Quirk et al. 1985: 1278), e.g. *the occupants of the cars – de som satt i bilene* ('those who sat in the cars'). Whereas it may be hard to find a Norwegian noun phrase corresponding idiomatically (in this context) to *the direction of the voice* in (22), other examples may simply show a preference for clausal rather than phrasal expression, e.g. (23), where the Norwegian verb *hete* ('be called') is available to convey the meaning of *the name of*.[8]

(23)   *The name of the gallery* is Sub-Versions, one of those puns that used to delight me before they became so fashionable. (MA1)
       Galleriet heter Sub-Versjoner, en type ordspill som pleide å more meg før det gikk inflasjon i dem. (MA1T)
       Lit: "Gallery$_{def}$ is-called Sub-Versions…"

5.5 Variation across particular N1s
This final part of the corpus investigation turns to the most frequent lexical items to be found in the N1 position in the colligational framework under study: *middle, edge, rest,* and *end*. As shown in Figure 2, the four lexemes vary greatly as regards their typical correspondence

---

[8] The Norwegian verb *å hete* does not have a straightforward English equivalent, but means 'to be called' (cf. French *s'appeller* and German *heissen*).

patterns. *Middle* has mostly divergent correspondences in both translations and sources; *edge* has more divergent sources than translations; *Rest* has the greatest proportion of congruence of the set, and *end* shows opposite trends in sources and translations.
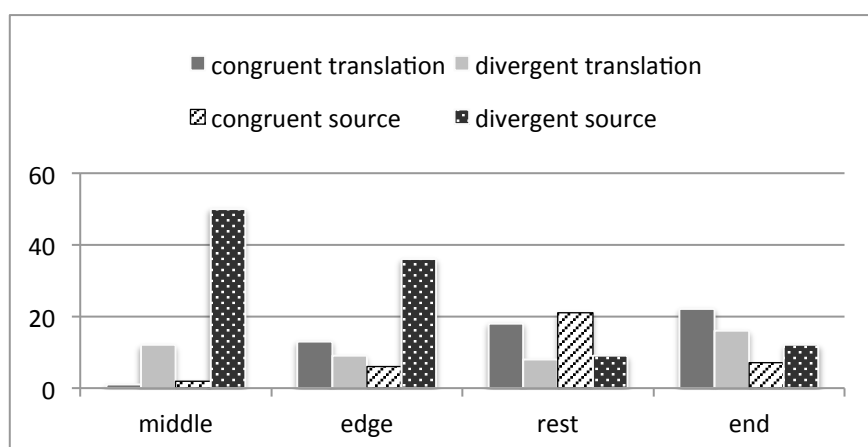

Figure 2. Congruent and divergent correspondences across the most frequent N1s (raw frequencies).

As pointed out repeatedly below, the most frequent correspondence *par excellence* of *(in) the middle of the* is *midt i/på*. Although both *in the middle of* and *midt i/på* may be seen as complex prepositions, *in the middle of* also has a congruent correspondence in Norwegian, shown in (24). However, this expression is very infrequent, with one occurrence in Norwegian translations and two in Norwegian originals. Divergent correspondences other than *midt i/på* are also very infrequent.

(24)   The man *in the middle of the semicircle* was taller than the others. (MN1T)
Mannen *i midten av halvsirkelen* var høyere enn de andre. (MN1)
Lit: "Man$_{def}$ in middle$_{def}$ of semicircle$_{def}$ was taller than the others."

The edge of the N2 is another sequence with mainly divergent correspondences. The most frequent divergent correspondence type of *the edge of the* is compound, e.g. *bordkanten* ('the table-edge')*, bassengkanten* ('the pool-edge')*. Other divergent correspondence types

are rare, with N1 omission being the second most frequent type, exemplified by (25). As shown in Figure 2, *the edge of the* N2 also has congruent correspondences, illustrated by (3) above.

(25) Her long, grey braids were barely visible over *the edge of the skin rug*. (TTH1T)
Den lange, grå fletta hennes stakk såvidt fram fra *skinnfellen*. (TTH1)
Lit: "The long, grey braid$_{def}$ hers showed barely forth from skin-rug$_{def}$."

The rest of the + N2 differs from the two sequences discussed above by having predominantly congruent correspondences. The N2 in this sequence is often temporal, e.g. *resten av dagen / uken* ('the rest of the day/week'). Divergent correspondences are mainly non-recurrent except N1–adjective in Norwegian translations, as shown in (26). The adjectives corresponding to *rest* as N1 in this pattern are quantifying; thus they reflect the quantifier character of the sequence *the rest of*.[9]

(26) She had completely forgotten now about *the rest of the class*. (RD1)
Hun hadde fullstendig glemt alle *de andre elevene*. (RD1T)
Lit: "She had completely forgotten all the other pupils$_{def}$."

The last sequence to be discussed in this section, *the end of the,* is congruent more often in translation and divergent more often in sources. However, the difference is small. The most frequent type of divergence found in translation is clausal expansion (six cases), exemplified by (27). In sources the most frequent divergences are *s*-genitive (four examples), as in *the end of the world – verdens ende* ('the world's end')*,* and N1 corresponding to an adjective, as in (28).

(27) Now, near *the end of the day*, it was clear that hope had been in vain. (AH1)

---

[9] In this respect *the rest of the* is analogous to other quantifying phrases with *of* such as *a lot of* (cf. Renouf & Sinclair 1991), which may be the clearest cases of the N2 acting as notional head of the noun phrase (cf. Owen 2007: 212).

Nå da *dagen nærmet seg slutten*, var det tydelig at håpet hadde vært forgjeves. (AH1T)
Lit: "Now when day$_{def}$ neared itself end$_{def}$, was it clear…"

(28)   At *the end of the working day*, with a more or less favorable outcome in my briefcase, it feels good to step into the dim interior of the restaurant… (KF1T)
Etter *endt arbeidsdag* med et mer eller mindre lovende resultat i dokumentmappen, føles det godt å trå inn i den halvmørke restauranten… (KF1)
Lit: "After ended workday with a more or less promising result in briefcase$_{def}$…"

6 Summary of findings and concluding remarks
The present study has been exploratory in that it has tested the possibility of using a colligational framework as the starting point for contrastive analysis. A number of findings have been presented. First of all, "the N1 of the N2" seems to select the N1 more systematically than the N2. The selected N1s typically focus on a part of the N2 (cf. Sinclair 1991), often with locative/temporal meaning. Possessive meaning, expressed by *of*-genitives, is also frequent. Support nouns, however, which were common in the pattern "a _ of" (Renouf & Sinclair 1991), are not particularly frequent in the present material.

   Norwegian correspondences of the framework are diverse, and differ between the directions of translation. Congruent correspondences, defined as "definite noun + preposition + definite noun", account for 44% of Norwegian translations and 33% of sources, and the difference is greater if the pattern is part of a prepositional phrase (section 5.2). Of the structural types of divergence, compound nouns were most frequent, closely followed by N1–adverb correspondences and *s*-genitives. It is noteworthy that most of the divergent correspondence types avoid the postmodification inherent in "the N1 of the N2" framework. This is in line with Elsness's (2014) findings that Norwegian uses fewer (clausal) postmodifiers than English, and that the content of the modification is more frequently expressed outside the noun phrase in Norwegian (2014: 117).

Some of the structural types of divergent correspondences differ across sources and translations: compounds and N1–adverb are more common in Norwegian sources; N1 omission and clausal correspondence are more common in Norwegian translations. The Norwegian preference for adverbs instead of nouns to denote location seems to be relatively systematic, and agrees with the general hypothesis that Norwegian prefers verbal to nominal style to a greater extent than English (cf. Nordrum 2007).

The broad meaning categories of the framework "the N1 of the N2" has also proved important for the preferred type of Norwegian correspondence. The most striking example is the category of locative N1s, which very often correspond to Norwegian adverbs, particularly if they are preceded by a preposition (in which case the Norwegian adverb corresponds to preposition + N1, e.g. *at the foot – nederst* ('nethermost')). Sequences where the N1 is a part of N2 can have congruent correspondences in Norwegian; however, it seems that Norwegian originals prefer other ways of expressing this relation, particularly compound nouns. It should be noted that Norwegian has a prepositional genitive analogous to the *of*-genitive, which accounts for the comparative frequency of congruent correspondences in the categories part-whole and owned/caused by N2.

As shown in Section 5.5, individual lexemes may also come with their preferred correspondence types: (*in*) *the middle of* regularly corresponds to *midt i/på*; a correspondence that seems quite conventionalized and is found e.g. in the bilingual dictionary *Engelsk stor ordbok* (Kunnskapsforlaget 2001). Another frequent N1, *edge*, regularly corresponds to the second part of a compound noun. *The rest of* corresponds congruently and frequently to *resten av*, while *the end of* has less clear correspondence patterns. A less frequent sequence, *the name of*, was shown to correspond to a clausal expansion with the verb *hete*, thus showing how languages may possess different lexicogrammatical resources for expressing the same meaning.

This investigation has been exploratory, and a number of issues remain for future studies, such as investigations of related colligational patterns (e.g. "a(n) N1 of a(n) N2") and their Norwegian correspondences. There is good reason to believe that such colligational frameworks are register-sensitive, cf. Biber et al (1999: 1014); thus a study involving more registers would also be worthwhile.

A potential problem of the present study is that the search was unidirectional, going only from English to Norwegian. This is probably unavoidable in an investigation based on function words: the languages cannot be expected to have function words that are similar enough to perform the same functions in phraseological frameworks. In this case, the main problem was the morphological difference in how definiteness is marked, with the Norwegian definite suffix even varying for gender and number. In addition, there is probably no Norwegian preposition with the same degree of multifunctionality as English *of*, so there is no way of searching lexically for Norwegian sequences equivalent to "the N1 of the N2". The tagged ENPC allows searches for two nouns with an intervening preposition, but for the present framework, this search string has a very low degree of precision; in most of the hits, the PP was not part of the same phrase as the N1, and due to the lack of morphological annotation, the search string cannot discriminate between definite and indefinite nouns. Furthermore, with the present extent of corpus annotation it is impossible to search for Norwegian compound nouns, which might have been an interesting starting point of a different investigation. There may thus be a need for developing more sophisticated ways of searching for patterns across languages.

In spite of its limitations, the present study has confirmed that English and Norwegian differ in their use of noun phrases. The differences are to some extent structural, but more importantly, the languages seem to select differently from relatively similar lexico-grammatical resources: Norwegian uses compound nouns more than English does, and favours a number of expressions that are structurally less complex than the English "the N1 of the N2" framework.

The most important implication of this study is arguably its demonstration of the feasibility of exploring structures and meanings cross-linguistically on the basis of function words. Such a way of exploiting parallel corpora represents a bottom-up approach to the relationship between form(s) and meaning(s) and can "give new insights into the languages compared – insights that are likely to be unnoticed in studies of monolingual corpora" (Aijmer & Altenberg 1996: 12). Renouf & Sinclair observe that the study of collocational frameworks can "raise consciousness of the many different and eminently sensible ways we might develop to present and explain language patterning" (1991: 143). This is certainly the case in cross-linguistic study too. Let this final

corpus example illustrate how and why different languages have different preferences in expressing similar meanings:

(29)   It's the way of the world. (RD1)
       Sånn er det bare. ("Such is it simply")

References
Aijmer, Karin & Bengt Altenberg. 1996. Introduction. In K. Aijmer, B. Altenberg & M. Johansson (eds.), *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*. Lund: Lund University Press, 11-16.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Brems, Lieselotte. 2015. Some binominal size noun constructions in English and French: a contrastive corpus-based perspective. *Nordic Journal of English Studies* 14:1, 90-115.

Elsness, Johan. 2014. Clausal modifiers in noun phrases: a comparison of English and Norwegian based on the Oslo Multilingual Corpus. S.O. Ebeling, A. Grønn, K. R. Hauge & D. Santos (eds), *Corpus-based Studies in Contrastive Linguistics*. Oslo Studies in Language 6(1), 91-118.

Engelsk stor ordbok. Engelsk-norsk / norsk-engelsk. 2001. Oslo: Kunnskapsforlaget.

Faarlund, Jan Terje, Svein Lie, Kjell Ivar Vannebo. 1997. *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.

Groom, Nicholas. 2007. *Phraseology and Epistemology in Humanities Writing*. PhD thesis, University of Birmingham. Accessed from http://ethos.bl.uk, 30 July 2015).

Hunston, Susan. 2008. Starting with the small words. Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics* 13:3, 271-295.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies*. Amsterdam: Benjamins.

Johansson, Stig. 2012. Cross-linguistic perspectives. In Merja Kytö (ed.), *English Corpus Linguistics: Crossing Paths*. Amsterdam / New York: Rodopi, 45-68.

Johansson, Stig, Jarle Ebeling & Signe Oksefjell. 1999/2002. English-Norwegian Parallel Corpus. Manual. Department of British and American Studies, University of Oslo. http://www.hf.uio.no/ilos/tjenester/kunnskap/sprak/omc/enpc/ENPCmanual.html, accessed 29 July 2015.

Keizer, Evelien. 2007. *The English Noun Phrase*. *The Nature of Linguistic Categorization.* Cambridge: Cambridge University Press.

Miccoli, Annalisa. 2010. "'We can pull it off, if we hold on... Men skal det ta av, hvis vi står på?' *Off* versus *av* and *on* versus *på*: a corpus-based contrastive investigation". MA thesis, University of Oslo. Available at http://urn.nb.no/URN:NBN:no-25295.

Nordrum, Lene. 2007. *English Lexical Nominalizations in a Norwegian-Swedish Contrastive Perspective*. PhD thesis, Göteborg University. Available at https://gupea.ub.gu.se/handle/2077/17181.

Owen, Charles. 2007. Notes on the ofness of *of* – Sinclair and grammar. *International Journal of Corpus Linguistics* 12:2, 201-222.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Renouf, Antoinette and John Sinclair. 1991. Collocational frameworks in English. In K. Aijmer and B. Altenberg (eds). *English Corpus Linguistics: Studies in Honour of Jan Svartvik.* London: Longman, 128-144.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Corpora

English-Norwegian Parallel Corpus, see http://www.hf.uio.no/ilos/tjenester/kunnskap/sprak/omc/enpc/

British National Corpus, see http://www.natcorp.ox.ac.uk/

Leksikografisk bokmålskorpus, see http://www.hf.uio.no/iln/tjenester/kunnskap/samlinger/bokmal/veiledningkorpus/

Contrasting translations

Thomas Egan, Hedmark University of Applied Sciences

Abstract
This paper argues that corpora containing translations of the same source text into two or more target languages represent a sounder base for comparing linguistic expressions in several languages than do corpora containing data from just two languages, whether they be unidirectional or bidirectional. The argument is illustrated with comparisons of some prepositional constructions in English and French using data from the Nor. – Eng. – Fr. – Ger. part of the Oslo Multilingual Corpus, and some constructions in Swedish and Norwegian using the overlapping original texts in the ENPC and ESPC.

1. Introduction

In the last twenty years, since the widespread availability of translation and parallel corpora, much productive work has been done comparing translations with original texts, both the original texts underlying the translations themselves, and other texts in the target language which are employed to check the extent to which the translations are representative of usage in that language. Indeed there have been over 200 such publications based on the corpora whose anniversary is being celebrated in this special edition. A good deal less work has been done on comparing translations into different languages of one and the same set of source texts. Some examples of the latter approach are Johansson (2001), Paulussen (1999), Slobin (2005), Verkerk (2014, 2015) and Viberg (1998, 2003, 2013).

In this paper I explore the advantages of comparing translations into different languages of the same source texts, rather than original texts and their translations, illustrating my argument with comparisons of some constructions in English and French with data from the Norwegian – English – French – German part of the Oslo Multilingual Corpus (see Johansson 2007, Egan 2013a, 2014, 2015a, 2015b)),[1] and some

---

[1] For the sake of convenience the abbreviation OMC will hereafter be used for the Norwegian – English – French – German part of the Oslo Multilingual Corpus (<https://www.hf.uio.no/ilos/english/services/omc/>) even though, strictly speaking, the term OMC denotes a whole family of corpora.

constructions in Swedish and Norwegian using the overlapping original texts in the English-Norwegian Parallel Corpus (ENPC) and the English-Swedish Parallel Corpus (ESPC) (see Hasselgård 2007, Rawoens & Egan 2015).

In section 2 I describe what I see as the weaknesses inherent in using original texts and their translations to compare expressions in two languages. Section 3 contains a comparison of the use of visual and verbal prompts in the elicitation of predications of motion for the purpose of contrasting these in two or more languages. In section 4 I present some sorts of predications which cannot be elicited using visual prompts. Section 5 contains details of the overlapping texts in the ENPC and ESPC and argues that these texts comprise a viable corpus for contrastive studies of expressions in Norwegian and Swedish. Finally, section 6 contains a summary and conclusion.

2. Comparing original texts and translations

When comparing translations with original texts, one is faced with the problem that these are different text types, one of which may display translation effects, the other not. By translation effects (Johansson 1998: 5), or translationese (Gellerstam 1996: 54), are meant the retention in the target language texts of features of the source language that are not equally felicitous in the target. So prevalent are these features that, as has been demonstrated by Cappelle (2012), it is sometimes possible to use them to predict the original language of a translated text. This problem was recognised by the compilers of corpora such as the ENPC and ESPC and was addressed in them by including texts and translations in both directions, thus enabling the comparison of original and translated language in the two languages being compared (see Johansson 2007).

Any comparison between two or more features is dependent on the availability of a viable *tertium comparationis*. According to Johansson (2001: 584), "The advantage of a corpus of original texts and their translations is that the translation is intended to express the same meaning as the original text". However, Altenberg & Granger, writing just a year after Johansson, maintain that:

> It is not realistic to proceed from a *tertium comparationis* that is based on 'identity of meaning'. For one thing, this would be putting the cart before the horse and we would run the risk of methodological circularity: the result of the contrastive

analysis would be no more than the initial assumption (cf. Krzeszowski 1990: 20). For another, the area we want to explore is often fuzzy and impossible to define satisfactorily. (Altenberg & Granger 2002: 16)

Krzeszowski' s original formulation of the first objection reads:

> We compare in order to see what is similar and what is different in the compared materials; we can only compare items that are in some respects similar, but we cannot use similarity as an independent criterion in deciding how to match items for comparison since similarity (or difference) is to result from the comparison and not to motivate it. (Krzeszowski 1990: 20)

However, one could argue that this objection is not equally trenchant if the similarity that is originally postulated ('in some respects similar') is broader, more coarse-grained than the similarity/difference that is the goal of the study. The analysis would then resemble a spiral more than a circle, to pursue the geometric metaphor, and although we follow a curved path in negotiating a spiral, we do not arrive back at our original destination.

The second objection raised by Altenberg and Granger carries, I think, more weight. It is often difficult to formulate even a satisfactory coarse-grained definition of the research object. Ebeling & Ebeling (2013: 21) make the same point, contending that "One of the difficulties in starting with meaning is how to delimit it. Starting with form, the boundaries are already set, while meaning is much more elastic."

Related to the difficulty of pinning down meaning in two texts is a more basic problem involved in comparing originals and translations, which is that these two text types are produced subject to two different sorts of constraints. As researchers we only have access to one of these. That is, we know what the translator is trying to convey, but can only guess at the intentions of the author of the original text. The discrepancy is illustrated in Table 1, where the term '2-text corpora', borrowed from Krzeszowski (1990), is used for translation corpora containing texts in two languages.

Table 1. Sources and targets in 2-text translation corpora

	To be encoded	Encoded by
Translator	Expression in source text	Expression in target text
Original author	?	Expression in source text

We see in Table 1 that the expression in the source text occurs in two columns, in the second column as a prompt for the expression in the translated text and in the third as an utterance to be compared to the latter. The prompt in the third row, the content of which is represented by the question mark, is nebulous compared to its counterpart in the second row. Since the production of a meaningful utterance involves making a series of lexical and grammatical choices, it is vital for the analyst to be aware of the parameters within which these choices are made. However, in the case of the original author, as opposed to the translator, the analyst is in the dark as to the exact nature of the prompts in question, having to reason backwards from their expression in the third column.

The fact that 2-text translation corpora are less than ideal for the comparison of expressions in two languages should not be taken to mean that they cannot be mined for data about one of these languages: quite the contrary, in fact. A series of expressions in the target language may be used to shed light on the polysemanticity of a single item in the source language, for instance, as illustrated in Table 2.

Table 2. Translations in 2-text translation corpora of the same expression in the source language.

	To be encoded	**Encoded by**
Translator	Expression (a1) in source text	Expression (x) in target text
Translator	Expression (a2) in source text	Expression (x) in target text
Translator	Expression (a3) in source text	Expression (y) in target text
Translator	Expression (a4) in source text	Expression (z) in target text

Table 2 illustrates various translations of one and the same expression in one or more source texts, by one or more translators. It shows that two tokens of expression (a) are translated by expression (x), one by (y) and one by (z). With respect to translations of prepositions, the part of speech which I will be drawing on for the exemplification of my arguments in sections 3 – 5, Garretson (2004) maintained that "… if we take as our default assumption that similar forms will be used to translate similar meanings, we must expect that related meanings will be expressed with the same form more often than unrelated meanings will" (Garretson 2004: 23). The method of analysis illustrated in Table 2 has been employed in studies based on the ENPC and ESPC to throw light on a large variety of linguistic items (for just one of many possible examples, see Aijmer, this volume). The greater the number of source texts, and the

greater the number of translators, the more light the analyst can shed on the expression in the original language as reflected in the various translations. However, if one is interested in comparing expressions in two or more languages, rather than in using translations as a means to explore the polysemanticity of items in a source language, 2-text corpora cannot furnish us with optimal *tertia comparationis*, for reasons outlined above. In the next section I present a couple of alternative sources of data and argue that these do not suffer from the same shortcomings.

3. Contrastive research into spatial relations

Given the problems inherent in the comparison of an expression in an original text with one in a translated text described in the previous section, one might ask whether there are any alternative *tertia comparationis* that the researcher may have recourse to. One way of ensuring that the linguistic items being compared are produced under similar constraints is to provide informants with a *tertium comparationis* from another modality. In research into predications of location and motion, this *tertium comparationis* may take the form of drawings, picture books or video snippets (see, for instance, Engemann *et al*. 2012, Vulchanova *et al*. 2012). The most extensively used text in the elicitation of descriptions of motion in various languages is the *Frog where are you?* story (Berman & Slobin 1994). Figure 1 contains an illustration from this story showing the frog climbing out of its jar.

Figure 1. A picture from the *Frog where are you?* story (Mayer 1969)

Table 3 summarises the process and results of an elicitation experiment, such as that carried out by researchers using visual prompts.

Table 3. Sources and targets in picture-based elicitation experiments

	To be encoded	Encoded by
Informant 1	Picture (a) in source text	Expression in first target text
Informant 2	Picture (a) in source text	Expression in second target text

Comparing Table 3 to Table 1, we see that in Table 3 the constraints on the informants are identical, insofar as both are responding to the exact same stimulus. Moreover, this is a stimulus to which researchers have full access.

Instead of using pictures, still or moving, as prompts, we can use verbal texts. A set of source texts and two sets of target texts make up

what we may call a '3-text corpus' (see Egan 2013a). Informants are still provided with identical prompts, but the *tertia comparationis* are verbal rather than visual. It may be objected that the informants, in this case translators, are subject to more constraints than observers of a picture, insofar as they may not be at liberty to construe the situations or events in the original texts as freely as the latter. They are, however, at liberty to *re*-construe them, should they feel the wish or need to do so. This point may be exemplified by (1) and (2), which consist of [EXIT] predications, similar to that illustrated in Figure 1, produced by English and French translators in response to the Norwegian originals in the OMC (see Egan & Graedler 2015).

(1)   a. Jeg åler meg *ut av* vinduet igjen. (NF1)[2]
       b. Wriggling *through* the window … (NF1TE)
       c. Je me suis glissé à nouveau *par* la fenêtre. (NF1TF)

(2)   a. Hun holdt hesten an da hun var kommet *ut av* den siste kløfta.
       (HW2)
       b. When she rode *out of* the last crevice, she reined in her horse.
       (HW2TE)
       c. Elle retint le cheval après avoir *passé* le dernier ravin. (HW2TF)

In (1) both the English and the French translator code the manner of motion in the verb and the path in an adverbial, thus preserving the coding options of the original text. In (2) on the other hand, in which the original text contains a neutral verb of motion and a path adverbial, the English translator employs a manner motion verb and a path adverbial and the French translator the path verb *passer* 'pass'. Note that the inclusion of the Norwegian originals in the examples is not for the purpose of comparing them to the translations, but rather to illustrate the common prompts to which the translators are exposed. Table 4, which summarises the process and results of the translation endeavours, may be compared to Table 3, which it closely resembles, only differing in the nature of the prompts.

---

[2] The letters and number 'NF1' refer to the source text in the Norwegian – English – French – German sub-part of the Oslo Multilingual Corpus, 'NF' being the initials of the author. 'TE' and 'TF' stand for English and French translated text, respectively.

Table 4. Sources and targets in 3-text translation corpora

	To be encoded	Encoded by
Translator 1	Expression (a) in source text	Expression in first target text
Translator 2	Expression (a) in source text	Expression in second target text

There is, in fact, one other difference between the processes illustrated in Tables 3 and 4, which may not be immediately obvious. This is the fact that, unlike informants in visual experiments, translators represented in corpora such as the OMC are not participating in an experiment. There is therefore no danger of the results being contaminated by the observer's paradox (Labov 1972).

4. Advantages of using verbal as opposed to visual prompts

While the use of visual prompts in the elicitation of utterances about motion described in section 3 has led to a great increase in our knowledge of how these predications are coded in various languages, they are unfortunately only suited for the investigation of concrete relations, such as location and motion. For more abstract relations, we must turn to verbal prompts (see Egan 2015a). Such relations are moreover very common, as is shown in Table 5, where spatial predications account for just under 50% of the tokens of the Norwegian preposition *gjennom*, the default English translation of which is 'through' (see Egan 2014). To investigate the remaining tokens, we cannot have recourse to visual *tertia comparationis*.

Table 5. Types of 'throughness' encoded by *gjennom* in the ENPC (Egan 2014)

Semantic field	Number of tokens
Motion	146
Perception	62
Other (metaphorical)	46
Time	24
Idiom	13
Medium	8
Means	8
Location	6
Total	313

Leaving aside predications of motion and location, the three most common types of relationship coded by *through* are perceptual, temporal and other (metaphorical). This last category is a very heterogeneous one, consisting of tokens displaying a wide range of metaphorical mappings. The perceptual and temporal categories are more tight-knit, displaying less semantic variation in the source texts. Figures 2 and 3 illustrate the correspondences between English and French translation correspondences of *gjennom* in these two categories, divided in English into translations containing 'through', translations containing other prepositions and translations containing non-prepositional constructions, and in French translations containing the proposition 'à travers', translations containing other prepositions, translations containing the verb 'traverser' and translations containing other non-prepositional constructions.[3]
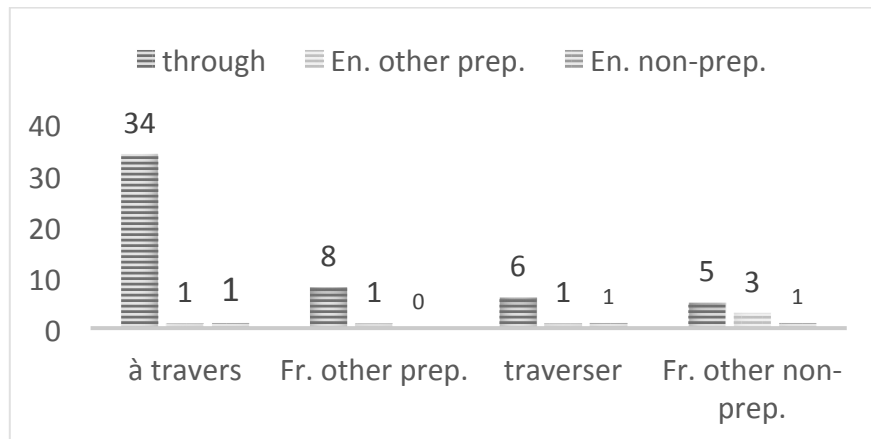
Figure 2. Correspondences between English and French codings of Perception [throughness] (Egan 2014)

In Figure 2, we can see that of a total of 62 tokens, 53 are translated into English by *through* and 36 into French by *à travers*. 34 of the latter

---

[3] The reason why the tokens are divided into four categories in French as opposed to just three in English, is the incidence in French of a large number of tokens containing the verb *traverser*. In a bottom-up investigation, such as this one, the categories must be allowed to emerge from the data, rather than being superimposed on the data in advance of the analysis.

correspond to English *through*. In one case a French preposition other than *through* corresponds to a similar English preposition and in three cases both sets of translators employ a non-prepositional construction. Thus the two sets of translators employ similar strategies in 38 cases (61%). Moreover *through* and *à travers* correspond to one another in codings of [throughness] involving various perceptual senses, such as vision in (3), hearing in (4) and smell in (5).

(3)   a. *Gjennom løvverket* skimtet jeg mesterens koie. (NF1)
       b. There was my master's hut visible *through the canopy of leaves*. (NF1TE)
       c. J'entrevois la baraque de Léopold *à travers le feuillage*. (NF1TF)

(4)   a. Måkene skrek inn til dem *gjennom* åpne vinduer. (HW2)
       b. Sea gulls shrieked to them *through the open windo*ws. (HW2TE)
       c. Les mouettes criaient *à travers les fenêtres ouvertes*. (HW2TF)

(5)   a. Han hadde teven av Dina tvers *gjennom* sildetønner …(HW2)
       b. He sensed Dina's aroma even *through the odors of herring barrels…* (HW2TE)
       c. L'odeur de Dina lui parvenait *à travers les tonneaux de harengs* ... (HW2TF)

Figure 3 contains details of the correspondences between the two sets of translations in the case of temporal predications.
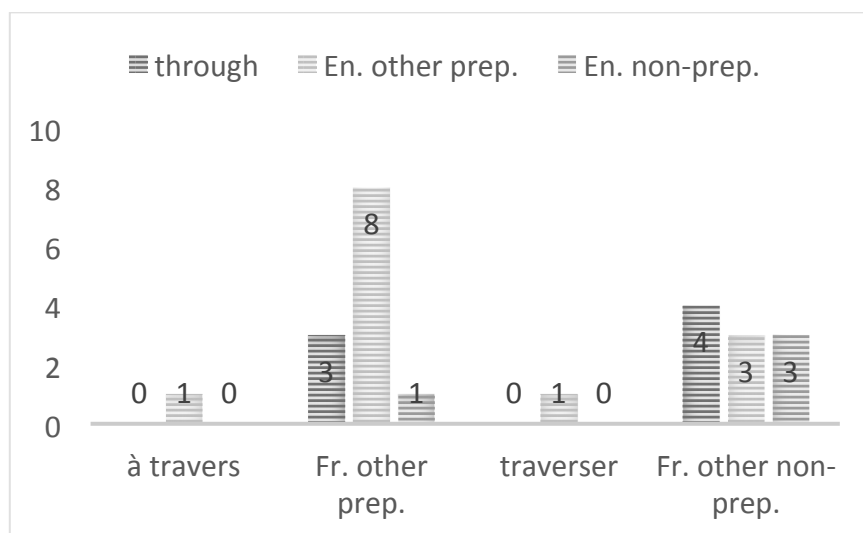
Figure 3. Correspondences between English and French codings of Time [throughness] (Egan 2014)

In Figure 3, we can see that translations of predications of temporal [throughness] differ markedly from those of perceptual [throughness] in that not so much as a single token is translated by both *through* and *à travers*. Of a total of 24 tokens just seven are translated into English by *through* and one into French by *à travers*. English employs alternative prepositions in 13 cases and French in 12. In eight cases both languages employ them, as in examples (6) – (7).

(6)　　a. Niels hadde brukt stor kløkt *gjennom flere år*. (HW2)
　　　　b. Niels had operated very cleverly *for many years*. (HW2TE)
　　　　c. Niels avait montré une grande intelligence *pendant plusieurs années*. (HW2TF)

(7)　　a. Trolig har han tatt for tunge løft i de unge år, for ut *gjennom manndomsalderen* gikk det så raskt tilbake med førligheten … (BHH1)
　　　　b. He probably overtaxed himself in his youth, for *in his mature years* his vigour declined so quickly … (BHH1TE)
　　　　c. Sans doute avait-il trop forcé au cours de ses jeunes années car, *durant sa maturité*, sa santé déclina si vite... (BHH1TF)

The French translators employ non-prepositional constructions in 11 cases, one of which contains the verb *traverser* and the English translators do so in four cases. Three of these overlap, including examples (8) and (9).

(8)   a. Men bildet dvergen knipset hadde allerede forfulgt Ana *gjennom* hele livet. (JG3)

          b. But the photo he'd taken had already hounded Ana *all* her life. (JG3TE)

          c. Mais ce portrait avait, en réalité, poursuivi Ana *toute* sa vie. (JG3TF)

(9)   a. 1. Han hadde vært medlem av Nasjonal Samling siden 1934 og betalt kontingent *gjennom* alle år. (BHH1)

          b. (1) He had been a member of the National Union since 1934 and paid his membership fee *every* year. (BHH1TE)

          c. 1) Il avait été membre du N.S. de 1934 à 1945 et s'était *régulièrement* acquitté de sa cotisation. (BHH1TF)

In (8) both translators employ a predeterminer rather than a preposition, *all* in English and *toute* in French, while in (9) the English translator employs the determiner *every* and the French one the adverb *régulièrement*, to code the fact that the actions in question occurred at regular intervals throughout the periods in question.

There is no doubt that the data presented in Figures 2 and 3 show that (translators into) English and French employ very different means of coding [throughness] with respect to perception and time. It is also clear that the sort of data presented in these two figures could not have been reliably elicited using pictures of any sort. Nor could they have been discovered using a 2-text corpus.[4] In any event it is hard to think of search items that would prove efficacious in unearthing them. And even were it possible to think of such items, the procedure for uncovering the relevant tokens would need to be very laborious. Yet the data are worth having insofar as they give us an indication of the correspondences

---

[4] An anonymous reviewer maintains that it would be possible to arrive at these conclusions using several 2-text corpora. While I agree that this might in theory be possible, one would still be faced with the problem of ensuring that the tokens in the two source corpora were as similar as possible.

between codings of certain perceptual and temporal predications in the two languages.

One type of relation that exhibits a great deal of overlap with respect to all sorts of predications is [betweenness], as is demonstrated in Egan (2013a) in a study of translations into English and French of the Norwegian preposition *mellom*. Table 6 contains details of the overlap between *between* and *entre* in coding the various sense of *mellom*.

Table 6. Tokens of *mellom* in the OMC translated by *between* and/or *entre* (Egan 2013a)

Type	Total tokens	between	entre	between + entre
Location	139	101	97	77 55%
Motion	89	50	60	40 45%
Interaction	51	44	36	32 63%
Relationship	47	43	35	33 70%
Comparison	36	33	23	21 58%
Time	17	13	11	11 65%
Idiom	14	8	8	7 50%
Total	393	292	270	221 56%

Of a total of 393 tokens of Norwegian *mellom*, 74% are translated into English by *between* and 69% into French by *entre*. Moreover the two prepositions correspond to one another in 56% of all cases. One may legitimately question the usefulness of this calculation of the overlap between the two prepositions, limited as it is to tokens that translate Norwegian *mellom*. We can however also use the 3-text corpus to work out the degree of mutual correspondence between the English and French prepositions *between* and *entre* using a method based on Altenberg (1999). Altenberg's method involves the division of the total number of occurrences in target texts in a 2-text corpus of item *a* translating item *b* and vice versa by the total number of occurrences of both terms in the two sets of source texts. Multiplying the result of this calculation by 100 gives us the percentage overlap of the two items, which Altenberg labels their 'mutual correspondence'. We can adapt this method to 3-text corpora by replacing the total of *a* translating *b* and vice versa by the total number of mutual occurrences in the two translated texts of *a* and *b* multiplied by two. We have to multiply by two, since the correspondence is in both directions, i.e. we count the *a*s corresponding to *b*s plus the *b*s corresponding to *a*s. We then divide the result by the total number of tokens of both *a* and *b* in the two translated texts rather than the source

texts, which by definition do not contain either item. There are a total of 365 tokens of *between* in the English translations in the OMC and 477 of *entre* in the French translations. 242 of these occur in parallel translations. Using the above formula, their degree of mutual correspondence in the translations, which we may label MC_t, can be calculated as follows:

$$MC_t = \frac{(\text{overlap } between/entre)*2*100}{\text{total } between + \text{total } entre}$$

$$MC_t = \frac{(242)*2*100}{365 + 477} = 59.61\%$$

This figure of 59.61% may be compared to the total for mutual correspondence of the two items in translations of *mellom* which, according to Table 6, is 56%. These figures may, of course, also be compared to results of calculations using Altenberg's original formula on the correspondence of the two items in 2-text corpora.

I round off this section by summarising what I see as the advantages of 3-text corpora for a comparison of items in two languages. In the first place 3-text corpora afford us a sound(er) basis for comparing two texts since in each case we know both what both language users are aiming to convey, since they are both translators, and how they go about encoding this. Moreover, unlike in the case of experiments using visual prompts, there is no danger of the results being contaminated by the observer's paradox. Lastly, 3-text corpora allow us to calculate the overlap between expressions in two languages using a different method to that of pairs of 2-text corpora. The results of the two methods may be mutually informative.

5. The ENPC and ESPC as a 3-text corpus

The source tests in the ENPC and the ESPC are to a large extent (approx. 80%) the same. As pointed out by Hasselgård (2007):

> The English-Norwegian Parallel Corpus (ENPC) and the English-Swedish Parallel Corpus (ESPC) are sister corpora with common objectives, design criteria, and structures; cf. Aijmer *et al* (1996:79 f). Especially their fiction parts also share a good number of texts. One advantage of this is that it is possible not only to compare

English to Norwegian and Swedish respectively, but also to compare different
translations of the English originals with each other. (Hasselgård 2007, section 1)

By eliminating tokens from texts that are exclusive to one corpus or the
other, we can thus use the ENPC and the ESPC for contrastive studies of
Swedish and Norwegian, just as we can use the Norwegian – English –
French – German part of the OMC for contrastive studies of English,
French and German. Hasselgård (2007) deals with adverbs of frequency
and usuality. She concludes that the jury is still out on the suitability of
ENPC and the ESPC for the sort of contrastive study she conducted, at
least if one is searching for "lexical alternatives, expecting the
translations used in the ENPC to be automatically usable in Swedish or
those in the ESPC to work in Norwegian" (Hasselgård 2007: section 8).
However, she is more positive to the usefulness of the 3-text corpus for
comparing constructions.

Hasselgård's study is based on the 24 overlapping fictional texts in
the two corpora.[5] However, there are 39 texts in all that feature in both
corpora, 24 fiction and 15 non-fiction, and the latter are equally suitable
for inclusion in a contrastive study. The actual texts that overlap are:

- In the ENPC:
    - Fiction: all texts minus AB1, BC1, TH1, DL1, DL2 and
      MM1
    - Non-fiction: all texts minus HB1, ROB1, SJG1, ML1
      and LTLT1
- In the ESPC:
    - Fiction: all texts minus DLO1
    - Non-fiction: all texts minus CAOG1, RH1, RL1, JPM1,
      AS1, ABB1, AZ1, PHA1, STO1, and all Nobel lectures
      and parliamentary speeches

The complete overlapping corpus of both fiction and non-fiction was
used by Rawoens & Egan (2015) in order to compare the Norwegian
preposition *mellom* with its Swedish counterpart *mellan*. Figure 4 shows

---

[5] The full lists of texts in the two corpora may be consulted at
<http://www.sol.lu.se/engelska/corpus/corpus/espc.html> (ESPC) and
<https://www.hf.uio.no/ilos/english/services/omc/> (ENPC).
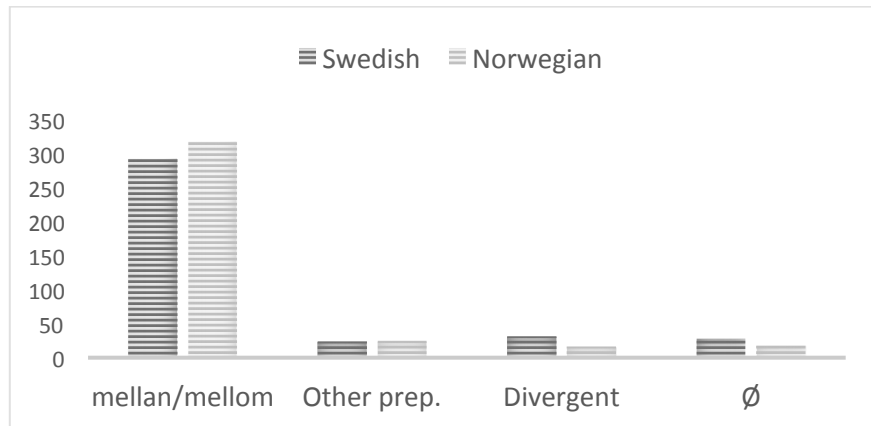
the total number of tokens of both prepositions when used to translate English *between*.

Figure 4. Swedish and Norwegian translation equivalents of English *between* (Rawoens & Egan 2015)

The overlap between the two prepositions in translations of *between* is 73%, while their degree of mutual translation correspondence (MC_t), calculated using the adjusted Altenberg method described in section 4, is 70%. There are more tokens in Norwegian than in Swedish (25 as opposed to 18) in which an alternative preposition is employed to encode a [betweenness] relationship. Especially common in Norwegian is the composite proposition *fra…til* ('from…to'), represented by 11 tokens. In nine cases these predications are rendered by *mellan* in Swedish, as in example (10).

(10) a. Some double stars are so close that they touch, and starstuff
      flows *between* them. (CSA1)
      b. Vissa dubbelstjärnor ligger så nära varandra, att de snuddar
      varandra och stjärnmateria flödar *mellan* dem. (CSA1TS)[6]

---

[6] 'TS' stands for Swedish translated text and 'TN' for Norwegian translated text, in accordance with the system for identifying texts in the OMC and utilised in examples (1) – (9).

c. Enkelte dobbeltstjerner står så nær hverandre at de berører hverandre, og stjernestoffet bølger *fra* den ene *til* den andre og tilbake igjen. (CSA1TN)

There are just over twice as many divergent constructions in the Swedish than the Norwegian translations of *between* (30 compared to 14). Seven of these are divergent in both languages. One of these is reproduced here as example (11). The Swedish translator has employed an adverbial meaning 'together' and the Norwegian translator a verb meaning 'help' and a reciprocal pronoun meaning 'one another'.

(11) a. She and Ryan *between* them cope with the baby. (PDJ1)
    b. Och *tillsammans* klarar hon och Ryan av babyn. (PDJ1TS)
    c. Hun og Ryan hjelper *hverandre* med den lille. (PDJ1TN)

Turning to tokens of *mellan* and *mellom* that do not translate *between*, there are 128 of these in the Swedish translations and 125 in the Norwegian translations. In 32 cases *mellan* and *mellom* overlap. One could argue that this overlap represents stronger evidence for the extent of semantic similarity between the Swedish and Norwegian prepositions than translations of *between*, since *mellan* og *mellom* must be considered the default translation correspondences of the English preposition. Two thirds of the tokens of *mellan* and *mellom* occur in congruent translations, with the most frequent original prepositions being *of*, *through*, *from*, *in* and *among*. In (12) both translators have chosen to translate *through* by their equivalent of *between*.

(12) a. … and picking their way with distaste *through* the puddles on the pavement. (PM1)
    b. De kryssar med avsmak *mellan* vattenpölarna på trottoarerna. (PM1TS)
    c. Så tar de seg forsiktig frem *mellom* søledammene på fortauet med sterkt misbilligende ansiktsuttrykk. (PM1TN)

Had the two translators chosen to employ *genom/gjennom*, the most common Swedish and Norwegian correspondences of *through*, the target texts would have differed from the source text in predicating that the

subjects actually put their feet in the puddles rather than wove their way around them.[7]

In one third of the cases where *mellan* and *mellom* cooccur but do not translate *between*, the source text contains a divergent construction, that is one without a preposition. One English construction that stands out as being avoided by both the Swedish and Norwegian translators is the coding of bilateral relations by means of a hyphen, as in 'East-West suspicions' in (13). Other original tokens of the same type are 'The Iran-Iraq war' (CS1), 'the Sino-Soviet split' (MAW1) and 'U.S.-British relations' (AH1).

(13) a. *East-West* suspicions have erected a visible barrier…(CS1)
      b. Den ömsesidiga misstänksamheten *mellan* öst och väst har
         upprättat en synlig barriär… (CS1TS)
      c. Mistenksomhet *mellom* øst og vest har reist en synbar skranke…
         (CS1TN)

In both (13b) and (13c) the suspicions are said to pertain *between* East and West. The coding of this relation by means of a preposition rather than a hyphen in both Swedish and Norwegian illustrates a feature common to these two languages. Such shared features are easy to spot in a 3-text corpus, since they stand out like individual trees in a wood. In a 2-text corpus consisting of texts from closely related languages, such as Swedish and Norwegian, on the other hand, their very similarity renders their identification more difficult.

6. Summary and conclusion
In this paper I have argued that 3-text translation corpora afford us possibilities for conducting contrastive research that are different to, and in some respects superior to, those afforded by (pairs of) 2-text corpora. Contrasting translations is different to contrasting original texts and translations insofar as the two items being compared have been produced in response to the exact same set of prompts, which function as the *tertium comparationis*. It is the identity of these prompts which led me to

---

[7] There are some similar examples of *mellom* translating *through* in Egan (2012: 48), including 'Diana picked her way *through* the women and answered her front door.' (ST1)

argue that they are also more suitable for comparison than a corpus consisting of original texts and translations of these, irrespective of whether the latter consists of translations in one or two directions. We have also seen that sets of translations in a 3-text corpus may furnish us with an alternative method for gauging the degree of mutual correspondence between linguistic items in the two languages concerned. This method is based on that proposed by Altenberg (1999). Finally, I hope I have shown that the common original texts in the ESPC and ENPC are suitable for contrastive analyses of expressions in Swedish and Norwegian.

References

Aijmer Karin, Bengt Altenberg and Mats Johansson, Eds. 1996. *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies Lund 4-5 March 1994*. Lund: Lund University Press

Altenberg, Bengt. 1999. "Adverbial connectors in English and Swedish: Semantic and lexical correspondences." *Out of Corpora. Studies in Honour of Stig Johansson*. Eds. Hilde Hasselgård and Signe Oksefjell. Amsterdam: Rodopi. 249-268.

Altenberg, Bengt and Sylviane Granger. 2002. "Recent trends in cross-linguistic lexical studies." *Lexis in Contrast: Corpus-based Approaches.* Eds. Bengt Altenberg and Sylviane Granger. Amsterdam: John Benjamins. 3-48.

Berman, Ruth and Dan Slobin. 1994. *Relating Events in Narrative: a Crosslinguistic Developmental study*. Hillsdale, NJ: Laurence Erlbaum.

Cappelle, Bart. 2012. "English is less rich in manner-of-motion verbs when translated from French." *Across Languages and Cultures* 13(2): 173-195.

Ebeling, Jarle and Signe Oksefjell Ebeling. 2013. *Patterns in Contrast.* Amsterdam: John Benjamins.

Egan, Thomas. 2012. "*Through* seen through the looking glass of translation equivalence: a proposed method for determining closeness of word senses." *Corpus Linguistics: Looking Back – Moving Forward*. Eds. Sebastian Hoffman, Paul Rayson and Geoffrey Leech. Amsterdam: Rodopi. 41-56.

Egan, Thomas. 2013a. "*Tertia Comparationis* in Multilingual Corpora." *Advances in Corpus-based Contrastive Linguistics. Studies in Honour of Stig Johansson*. Eds. Karin Aijmer and Bengt Altenberg. Amsterdam: John Benjamins. 7-24.

Egan, Thomas. 2013b. "Between and through revisited." *Corpus linguistics and variation in English: Focus on Non-native Englishes*, a special issue of the online-journal VARIENG: Studies in Variation, Contacts and Change in English. Eds. Magnus Huber and Joybrato Mukherjee.

Egan, Thomas. 2014. "Encoding *throughness* in English and French." *Diachrony and Synchrony in English Corpus Studies*. Eds. Alejandro Alcaraz Sintes and Salvador Valera Hernandez. Frankfurt: Peter Lang. 233-257.

Egan, Thomas. 2015a. "Manner and Path: evidence from a multilingual corpus." *CogniTextes* 12.

Egan, Thomas. 2015b. "Motion *to* and motion *through*: evidence from a multilingual corpus". *From Clerks to Corpora: Essays on the English Language Yesterday and Today*. Eds. Philip Shaw, Britt Erman, Gunnel Melchers and Peter Sundkvist. Stockholm: Stockholm University Press. 285-302.

Egan Thomas and Anne-Line Graedler. 2015. "Motion *into* and *out of* in English, French and Norwegian." *Nordic Journal of English Studies*, 14 (1): 9-23.

Engemann, Helen, Anne-Katharina Harr and Maya Hickmann. 2012. "Caused motion events across languages and learner types: A comparison of bilingual first and adult second language acquisition." *Space and Time in Languages and Cultures: Linguistic Diversity*. Eds. Luna Filipović and Katarzyna M. Jaszczolt. Amsterdam: John Benjamins. 263-287.

Garretson, Gregory. 2004. "The meanings of English "of": uncovering semantic distinctions via a translation corpus." Unpublished MA thesis, Boston University.

Gellerstam, Martin. 1996. "Translations as a source for cross-linguistic studies." *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies, Lund, 4–5 March 1994*. Eds. Karin Aijmer, Bengt Altenberg and Mats Johansson. Lund: Lund University Press. 53-62.

Hasselgård, Hilde. 2007. "Using the ENPC and the ESPC as a Parallel Translation Corpus: Adverbs of Frequency and Usuality." *Nordic Journal of English Studies,* 6(1).

Johansson, Stig. 1998. "On the role of corpora in cross-linguistic research." *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Eds. Stig Johansson and Signe Oksefjell. Amsterdam: Rodopi. 3-24.

Johansson, Stig. 2001. "The German and Norwegian correspondences to the English construction type *that's what*." *Linguistics* 39 (3): 583-605.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.

Krzeszowski, Tomasz P. 1990. *Contrasting Languages: The Scope of Contrastive Linguistics*. Berlin: Mouton de Gruyter.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.

Mayer, Mercel. 1969. *Frog, where are you?* New York: Dial Books.

Paulussen, Hans. 1999. "A corpus-based contrastive analysis of English *on/up*, Dutch *op* and French *sur* within a cognitive framework." Unpublished doctoral dissertation, Ghent University.

Rawoens, Gudrun and Thomas Egan. 2015. "Coding *betweenness* in Swedish and Norwegian translations from English." *Meta: Journal des traducteurs* 60 (3).

Slobin, Dan. 2005. "Relating Narrative Events in Translation." *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman.* Eds. Dorit Diskin Ravid and Hava Bat-Zeev Shyldkrot. Dordrecht: Kluwer. 115-129.

Verkerk, Annemarie. 2014. "Where Alice fell into: Motion events from a parallel corpus." *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech.* Eds. Benedikt Szmrecsanyi and Bernhard Wälchli. Berlin: Walter de Gruyter. 324-354

Verkerk, Annemarie. 2015. "Where do all the motion verbs come from? The speed of development of manner verbs and path verbs in Indo-European." *Diachronica*, 32 (1): 69-104.

Viberg, Åke. 1998. "Contrasts in polysemy and differentiation: Running and putting in English and Swedish." *Corpora and Cross-linguistic*

Research: Theory, Method, and Case Studies. Eds. Stig Johansson and Signe Oksefjell. Amsterdam: Rodopi. 343-376.

Viberg, Åke. 2003. "The polysemy of the Swedish verb komma 'come': A view from translation corpora." *Meaning Through Language Contrast: Vol 2*. Eds. Katarzyna M. Jaszczolt and Ken Turner. Amsterdam: John Benjamins. 75-105.

Viberg, Åke. 2013. "Seeing the lexical profile of Swedish through multilingual corpora: The case of Swedish *åka* and other vehicle verbs." *Advances in Corpus-based Contrastive Linguistics. Studies in Honour of Stig Johansson*. Eds. Karin Aijmer and Bengt Altenberg. Amsterdam: John Benjamins. 25-56.

Vulchanova, Mila, Liliana Martinez and Valentin Vulchanov. 2012. "Distinctions in the linguistic encoding of motion: evidence from a free naming task." *Motion Encoding in Language and Space*. Eds. Mila Vulchanova and Emile Van Der Zee. Oxford: Oxford University Press. 11-43.

Corpora
ENPC: http://www.hf.uio.no/ilos/english/services/omc/
ESPC: http://www.sol.lu.se/engelska/corpus/corpus/espc.html
OMC: http://www.hf.uio.no/ilos/english/services/omc/

What happens in translation? A comparison of original and translated texts containing verbs meaning SIT, STAND and LIE in the English-Swedish Parallel Corpus (ESPC)[1]

Åke Viberg, Uppsala University

Abstract
This article studies translation effects by comparing the use of verbs meaning SIT, STAND and LIE in original and translated texts in the English-Swedish Parallel Corpus (ESPC). Effects on both the frequency of use and the use of lexical and structural translation shifts are studied. Postural verbs have a much higher frequency overall in Swedish than in English. In Swedish translated texts, postural verbs are significantly under-represented in comparison to original texts, whereas postural verbs are significantly over-represented in English translations. At a more fine-grained level, it is possible to show that various categories are treated differently, in particular the types of subjects. The effect on frequency is stronger for Human subjects, which represent an unmarked category, than it is for Inanimate subjects, which are more marked. However, the pattern of over- and under-representation presupposes a functional overlap across languages. Writing as subject, which is a category that is unique to Swedish in relation to English, follows a different pattern. (This type of subject appears in examples such as: *Nyheten står i tidningen* 'The news is (literally: stands) in the paper'). The result is discussed both from the point of view of the research methodology used in contrastive studies based on translation corpora and from a theoretical point of view. For methodology, the conclusion is that frequencies can be considerably skewed, whereas a language remains true to its system of basic semantic contrasts in professional translations. Theoretically, the result can be related to theories of language contact and studies of second language acquisition and bilingual development.

1. Introduction: Contrastive studies and translation studies

1.1 Aim of the study

What happens in translation? Is the same content expressed in the translated text as in the original? Are translated texts different from original texts in the same language and, if so, to what extent is this difference dependent on the particular source language? In certain

---

situations of language contact, translated texts have been, and are, important forces behind language change – sometimes in conjunction with the influence from second language users of the target language. The present paper, focusing on postural verbs in original and translated texts in English and Swedish, will give only partial answers to some of these overarching questions, but it is important to keep the broader picture in mind. In addition to theoretical questions, an important aim of the paper is to test the methodology in contrastive studies based on translation corpora. The advantage of using translation corpora is that the expression of the same meaning in the same context can be compared across languages. On the other hand, the fact that translations do not always give a representative picture of original target language texts must be taken into consideration. In particular, it is important to identify which features are most vulnerable.

1.2 Why look at postural verbs?
The justification for looking at postural verbs in particular is that these verbs have been studied extensively from a typological (Newman 2002) as well as from a contrastive point of view (Lemmens 2002, Svensson 2005, Viberg 2013) and also from a second language acquisition perspective (Lemmens & Perres 2010). Of particular importance for the present study is the detailed and well-documented contrastive study by Kortteinen (2008) comparing the Swedish postural verbs *sitta* 'sit', *stå* 'stand' and *ligga* 'lie' with their French correspondents. That study contains a comparison of the use of the Swedish postural verbs in original and translated texts and shows that all three Swedish verbs are under-represented in translations.

In French, the salience of manner is low in descriptions of motion and location. The closest correspondents to SIT, STAND and LIE are not simple verbs in French and are used with a much lower frequency than in Swedish. English and Swedish, on the other hand, share historical cognates that still are very similar in form: *sit/sitta, stand/stå* and *lie/ligga*. The relationship between the postural verbs in English and Swedish represents a very special case that is likely to lead to maximal cross-linguistic influence. Semantically, not only the core meanings in reference to human posture are similar, but there is also a semantic overlap in several cases when postural verbs refer to the location of

physical objects (Viberg 2013). From a methodological point of view, this represents a suitable area for testing the validity of translation-based contrastive studies.

1.3 Earlier research

The answer to the question of what happens in translation can be approached from many different perspectives. The present study takes corpus-based contrastive studies (Johansson 2007) as a point of departure and confronts them with certain aspects of translation studies (Gambier & van Doorslaer 2010, Malmkjær & Windle 2011, Munday 2008, Halverson 2003, 2010) and in particular corpus-based translation studies (Laviosa 2002). From a contrastive point of view, this study approaches what Johansson (2007, 32-33) refers to as translation effects, i.e features which in some way make translated texts different from original texts in the same language. Johansson's term is neutral in comparison to the more value-laden term *translationese* used in Gellerstam's (1986) seminal article. An important issue is how to cope with translation effects from a methodological point of view in contrastive studies based on multilingual translation corpora such as Viberg (2013).

One specific type of translation effect, which has been much discussed in translation studies, is *translation universals* (Baker 1993, Mauranen & Kujamäki 2004), features that are characteristic of translated texts in general, irrespective of the source language. Examples of possible such effects are normalization and explicitation (see Malmkjaer 2011 for a balanced review). Typically, studies of translation universals are based on comparison between originals and translations of the same language. The present study will primarily be concerned with translation effects characteristic of specific source-target language pairs and is closely related to studies of interference, transfer or cross-linguistic influence in bilingualism and second language acquisition research (see Jarvis & Pavlenko 2008 for a broad review, cf. Viberg 1998). In contact linguistics, translation effects are an important source of convergence and divergence between languages (see Braunmüller & House 2009 and, for translation, in particular Becher, House & Kranich 2009.)

As is often the case in corpus-based studies, this study treats translation as a collective process without taking individual translators into consideration (see Johansson 2007, 197-215 for an exception). This

is justified by the general importance of translations as a societal phenomenon. What happens in translation can also be studied from the perspective of the translator (the producer) in studies of the translation process, which are concerned with what happens in the translator's mind. Such research is based on methods such as think-aloud protocols, eye-tracking and keystroke logging (see Englund Dimitrova 2010 for a concise review). A corpus-based study can also – albeit more indirectly – shed light on the translation process seen from the translator's perspective, in particular in studies of translating strategies (Gambier 2010).

1.4. Translation patterns

In this paper, the closest correspondent to a word (or multi-word unit) that appears in a translated text will be referred to as *a translation* (with a corresponding plural form: translations). Broadly, translations will be divided into *direct translations* (the corresponding postural verb in this case), *lexical translation shifts*, when some other word is used, and *structural translation shifts* such as restructuring and structural reduction. In addition, there are *free translations*, where the translation differs so much that it is impossible to identify any systematic structural relationship with the original text. A further category is *no translation*, which refers to examples where an extensive text has not been translated and the translation is lacking for reasons that are not structurally motivated as in reductions. A systematic description of all translations of a word along these lines will be referred to as the *translation pattern* of the word.

1.5 The corpus and its composition

This study is based on the English-Swedish Parallel Corpus (ESPC) compiled by Altenberg and Aijmer (2000) according to the same principles as the English-Norwegian Parallel Corpus (see Johansson 2007, 9-21). The ESPC corpus consists of original printed texts in Swedish and English together with their translations. The texts are divided into two broad genres: Fiction and Non-fiction with several subcategories. The sizes of the various sub-corpora of the ESPC are shown in Table 1 (see Electronic sources for more information).

Table 1. The composition and size of *the English Swedish Parallel Corpus* (ESPC)

	EO	ST	SO	ET	Total
Fiction	340,745	346,649	308,160	333,375	1,328,929
Non-Fiction	364,648	344,131	353,303	413,500	1,475,582
Total	705,393	690,780	661,463	746,875	2,804,511

EO = English Originals; ST = Swedish Translations; SO = Swedish Originals; ET = English Translations.

The corpus allows comparisons of various types. It can be used as a comparable corpus to compare originals in both languages in a contrastive study (see Section 2.1). Contrastive studies often also look at the translation patterns of the linguistic elements as a way of getting a better grasp of the meaning of the words that have been singled out for a study, for example all English translations of Swedish *stå* 'stand'. This means that originals in one language are compared with translations in the other language. The ESPC can also be used for corpus-based translation studies, which involves in particular the language-internal comparison of originals in one language with translations into the same language. The corpus also allows comparison between two genres (or registers): Fiction and Non-Fiction. This is relevant also for the present study, since the distribution of uses (or meanings) of the postural verbs varies dramatically in certain respects between Fiction and Non-Fiction. For reasons of space this aspect will be treated briefly in Section 6.

The rest of the paper is structured as follows: Section 2 presents global comparisons, which means that various uses (or senses) of the postural verbs are not taken into consideration. Section 3 presents a semantic classification of grammatical subjects on which the following quantitative analysis is based. In Section 4, this classification is applied in a quantitative study of the distribution of direct translations across subject classes. Section 5 is devoted to a fine-grained descriptive comparison of lexical and structural translation shifts, where grammatical constructions are also taken into consideration. Section 6 looks at the distribution of the major uses across genres (represented by fiction and non-fiction) followed by conclusions and discussion in Section 7.

2. Global comparisons

2.1 Swedish and English originals

Table 2 compares the frequencies of the postural verbs in original texts. The result is also shown in Diagram 1. As can be observed, all of the Swedish verbs are considerably more frequent than their English correspondents.

Table 2. Frequency of SIT, STAND and LIE in Swedish vs. English original texts in ESPC

	Number of occurrences				Significance of difference
	Raw scores		Per 100 000 words		
	Sw.	Eng.	Sw.	Eng.	One-Tail
SIT	557	339	84	48	p<0.0001
STAND	939	301	142	43	p<0.0001
LIE	733	204	111	29	p<0.0001
Corpus size	661,463	705,393			

All of the differences are statistically significant (p< 0.0001). Statistical testing has been done with "The Significance of the Difference Between Two Independent Proportions" (Vassarstats, see electronic sources), which calculates the z-ratio for the significance of the difference between two independent proportions. This test as well as the following ones are based on raw scores, whereas the diagrams presented below show the frequencies per 100 000 words.

	SIT	STAND	LIE
Swedish originals	84	142	111
English originals	48	43	29
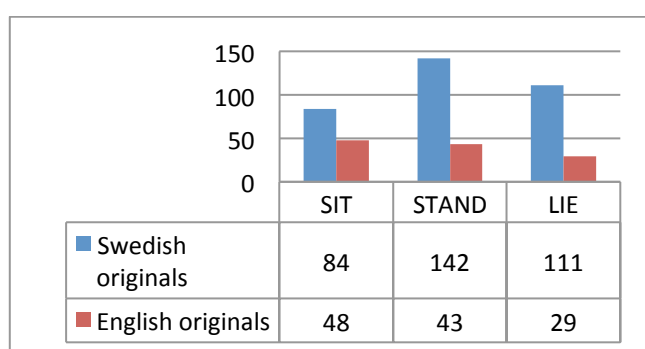
p<	.0001	.0001	.0001

Diagram 1. *Swedish and English Postural verbs.* The frequency in original texts of Swedish *sitta*, *stå* and *ligga* compared with English *sit*, *stand* and *lie.* Frequency per 100 000 words.

2.2 Intra-lingual comparison between originals and translations
Diagram 2ab shows what happens when the frequencies of the English and Swedish postural verbs are compared in original and translated texts.

	sit	stand	lie
Original	48	43	29
Translation	67	65	41
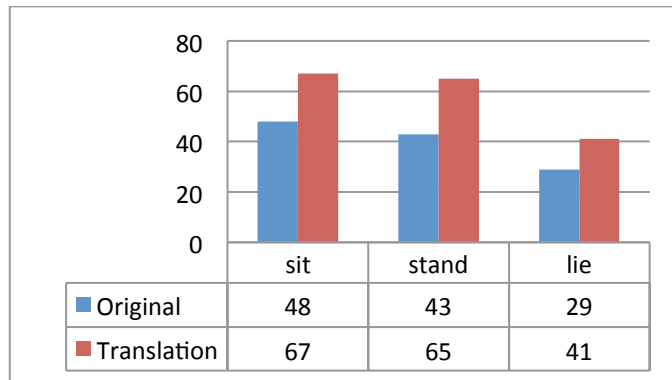
| p< | .0001 | .0001 | .0001 |

Diagram 2a. The Frequency of the English postural verbs: *sit, stand* and *lie* in originals and translations. Frequency per 100,000 words

In translated texts (Diagram 2a), English postural verbs are significantly over-represented (if original texts are regarded as the norm), whereas the Swedish equivalents are under-represented (Diagram 2b).

	sitta	stå	ligga
Original	84	142	111
Translation	69	105	92
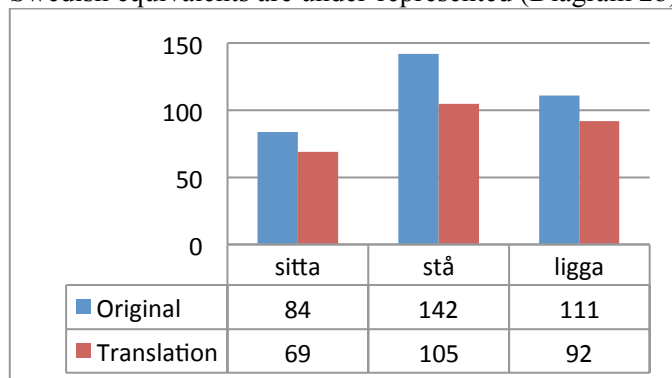
| p< | .0006 | .0001 | .0002 |

Diagram 2b The frequency of the Swedish postural verbs: *sitta stå* and *ligga* in originals and translations. Frequency per 100,000 words

As mentioned above, Kortteinen (2008) obtained a similar result as in Diagram 2b for Swedish texts translated from French. What is worth

noticing is that postural verbs are much more frequent and have a much wider range of uses in English than in French. Still, there is a strong translation effect. The results so far can be summed up as follows:

Contrastive relationships:
- In original texts, the Swedish postural verbs are much more frequent than their English direct correspondents in spite of the fact that all verbs are cognates which are transparently related in form and share core meaning.

Cross-linguistic influence:
- In translations from Swedish, English *sit, stand* and *lie* are significantly over-represented.
- In translations from English, Swedish *sitta, stå* and *ligga* are significantly under-represented.

At the global level, this points to a striking difference between originals and translations in the same language. However, this case cannot be generalized in a simple way. It turns out that other frequent lexical verbs do not always follow a similar pattern. The postural verbs were chosen as an example to test what happens in translation because they appear to represent a worst-case scenario from the point of view of contrastive studies based on translation corpora. (However, this applies only to mono-directional comparison. Bi-directional corpora like the ESPC can also be used to compare originals.)

From a general theoretical standpoint, it is interesting to get a more differentiated view of what happens. Are all uses (meanings) of a frequent polyfunctional word affected in the same way, or do we find translation effects only for some uses? What are the correspondents, when the postural verb does not correspond to a postural verb in the other language? The result so far seems to suggest severe translation effects, which does call for caution when using translated texts for contrastive comparison, but, as the results will reveal, a language stays true to its basic system of contrasts in professional translations and a comparison based exclusively on originals in comparable corpora misses important contrastive relationships. However, first a specific methodological problem must be addressed in the next section.

2.3 Variation due to genre: Fiction versus Non-Fiction
Variation due to genre (or register variation in general) is reflected in the use of individual words, but rather few of the many studies that, like the present one, deal with the polysemy of frequent words have tried to address this as a problem. Diagram 3ab shows the distribution of the postural verbs across the two genres of the ESPC.

	sit	stand	lie
■ Fiction	85	65	34
■ Non-Fiction	14	22	24

p<	.0001	.0001	.004
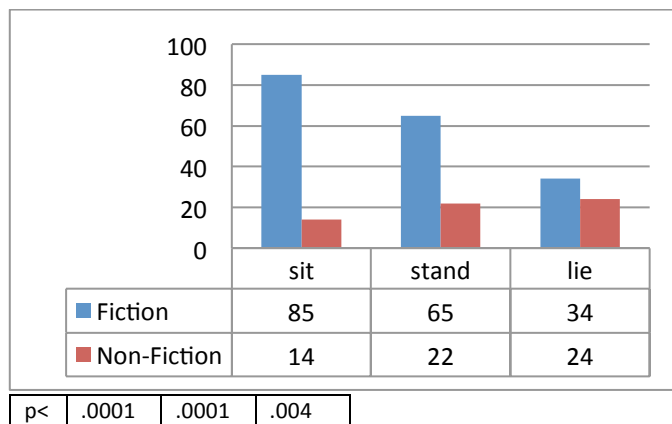
Diagram 3a. The English postural verbs *sit, stand* and *lie* in fiction and non-fiction. Frequency per 100,000 words

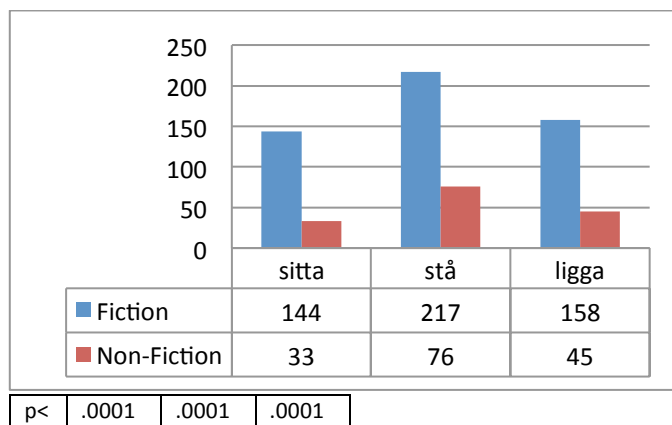|	sitta	stå	ligga
■ Fiction	144	217	158
■ Non-Fiction	33	76	45

p<	.0001	.0001	.0001

Diagram 3b. The Swedish postural verbs *sitta, stå* and *ligga* in fiction and non-fiction. Frequency per 100,000 words

There turns out to be very clear and significant differences. Both in English and Swedish, the postural verbs are considerably more frequent

in fiction than in non-fiction. The results are in line with the characterization of fiction as a distinct register (Ebeling & Ebeling 2013) and in particular with Stubbs & Barth's (2003, 65) conclusion that "verbs of physical movement, mental activity, and nouns for part of the body" are characteristic of fiction.[2] Non-fiction is negatively defined in the ESPC and even in actual practice covers a very broad variety of texts that serve as an interesting contrast to fiction, but need to be broken down into more clearly defined categories. (See, for example Biber et al. 1999, who compare fiction with conversation, news and academic prose.)

By comparing the frequencies in Diagram 3a and Diagram 3b, it is clear that each of the Swedish verbs is more frequent than its English counterpart in both fiction and non-fiction. This means that the conclusion reached in Section 2.1 concerning the contrastive relationship holds across registers (as far as we can tell from the ESPC corpus). As will be shown below, the register differences are more problematic when specific meanings or uses are taken into consideration, since certain uses are actually more frequent in non-fiction than in fiction. As a result, it is an open question whether fiction and non-fiction should be accounted for separately from the very beginning or not. In the coding of individual examples, the distinction is made, but the general account will present data from the whole corpus indiscriminately, and register differences will be discussed in Section 6.

3. Semantic classes of subjects

In order to test whether the global differences between originals and translations depend primarily on any specific meaning or are more evenly distributed across meanings, a functional-semantic coding of all examples was carried out. A natural starting point was the semantic category of the subject. The coding was done at different levels of granularity to make it possible to identify classes that were general enough to allow statistical comparison. At the coarsest level, four types of subject were distinguished: Human, Animal, Inanimate (concrete) and Abstract (see

---

[2] Postural verbs describe movement only in some uses but are closely related to such verbs. To be exact, we could refer to verbs of physical movement and posture as, perhaps, physical action in general.

Section 3.1). Different types of more fine-grained coding will be discussed in Section 3.2.

3.1 The four major classes

In its most frequent use, SIT, STAND and LIE have a **human** subject and refer to the posture of a human functioning as subject of the verb. Often, a locative complement is added. The Swedish verbs can also freely be used in presentative constructions (see (1)), which have the basic structure *det* 'it' + Verb + NP^{indefinite}. All the three Swedish postural verbs can appear in this structure with the same postural meaning, as in the corresponding non-presentative clause.[3]

(1)   *Det står* en man här ute som nödvändigtvis vill tala med dig. (HM2)
      [It stands a man]
      "*There's* a man out here who insists on talking to you." (HM2T)

In example (1), *a man* is the subject. To be exact, the classification is based on the **located entity** in cases like this one, but since the located entity in the majority of cases is realized as grammatical subject, this has been used as a general term in order not to complicate the terminology. In a few cases such as (2) and (3), when metonymy is involved, the choice of verb is decided by a category that is clearly different from the grammatical subject. In the Swedish original version there is a human subject, but the verb actually refers to the location of that person's car and boat, respectively. A similar type of metonymy is involved in: *I'm parked outside*. In example (3), *she* obviously refers to a boat in the English translation.

(2)   Kan en av er komma med mig och se om jag *står* bra där jag *står*. (SC1)
      [see if I stand all right where I stand]

---

[3] In all examples from the ESPC corpus, the original version appears first followed by a text code, e.g. (HM2) in (1), the reference of which is listed on the homepage of the ESPC project (see Electronic sources). Literal translations and other comments are sometimes given within square brackets. The original text is provided on top, followed by the translation. That's why there is a switch, for example in (4) below.

Could one of you come with me and check that I*'m* all right where I *am*? (SC1T)

(3)   Över vintern *låg jag i Dragör*, som hade en av de få hamnarna i Öresund där det var liv och rörelse året om. (BL1)
[Over the winter, I lay at Dragör]
At present, *she was laid up* for the winter at Dragor, that being one of the few ports on the Sound which remained astir whatever the season. (BL1T)

Such examples are coded as inanimate subject, since that is what decides the choice of verb. (In canonical position, cars stand and boats lie in Swedish.) There are not very many examples of this type, but they are theoretically interesting.

Animal subjects were so few that it is not meaningful to include them in most of the comparisons. It is tempting to collapse this class with human into an animate class, but the naming of animal postures follows different rules and certain animals have a default verb (birds sit in trees, snakes lie in the grass etc.).

Both English and Swedish can use the postural verbs to denote the location of **inanimate** concrete subjects. When the verbs meaning STAND and LIE are used, the basic semantic distinctions are similar, but there is a rather strong tendency to express location with the copula in English, whereas such uses are very restricted in Swedish. In both languages, LIE basically refers to an object placed in such a way that it has a maximal extension in the horizontal dimension, whereas STAND refers to a maximal vertical extension. In examples (4) and (5), the closest Swedish equivalent is used as a translation.

(4)   She came forward to open a folder that *lay* on the counter. (AT1)
Hon gick fram och slog upp en pärm som *låg* på disken. (AT1T)

(5)   The father took a golf-club that was *standing* in the corner. (RD1)
Hennes pappa tog en golfklubba som *stod* i ett hörn. (RD1T)

In Swedish, a special case is represented by physical objects (in particular artefacts such as a plate) that have a distinctive upper side, when they are used in a canonical way. The location of such objects is referred to with

stå, even when the horizontal extension is dominant (see example 6). Functional upper side appears to cover the same cases as *base* in other accounts (e.g. Lemmens 2002) but is based more on the typical human interaction with physical objects.

(6) Tallriken *stod* på en grå vaxduk med röd bård. (IB1)
     The plate *is* on grey oilcloth with a red border. (IB1T)
     [There is a tense shift in the translation from past to present.]

This special use is referred to as *functional upper side* in Table 3, which shows the basic contrasts in Swedish. Even when the subject is inanimate, the located entity is sometimes not explicitly referred to due to metonymy, for example, contents for container as in the Swedish original in (7), which literally means: 'It stood wine on the table'. The choice of verb in the Swedish example is based on the orientation of the bottle and is coded accordingly. In the English translation, the bottle is referred to explicitly.

(7) *Det stod* vin på bordet och en sallad.(KOB1)
     [it stood wine on table-the]
     There was a bottle of wine on the table, and a salad. (KOB1)

Table 3. The basic contrasts between Swedish Posture verbs indicating Inanimate Location.

Orientation	Posture verb
Horizontal dimension salient	*ligga* 'lie'
1.Vertical dimension salient 2. Functional upper side	*stå* 'stand'
Attached	*sitta* 'sit'

SIT represents a special case in both languages. According to Newman (2002:19), English *sit* basically has two functions with inanimate objects: (1) Non-activity *sit* "with a nuance suggesting that the entity is underutilized, out of use, useless etc." and (2) Good-fit *sit*, a position which represents a tidy arrangement or "good fit". Shape and orientation are irrelevant. An example of "good fit" *sit* is given in (8) and of "non-activity" *sit* in (9). In both cases, the translation is a verb that basically describes a different posture.

(8) (…) and one of the great attractions of this house was that it *sat*
within the boundaries of a National Park. (PM1)
(…) och något av det mest lockande med det här huset var att det
låg i en nationalpark. (PM1T)

(9) Their wedding picture used to *sit* on the piano in their living room.
(JSM1)
Deras bröllopsfoto brukade *stå* på pianot i deras vardagsrum.
(JSM1T)

In Swedish, *sitta* refers to an attachment of some kind. Both *stå* and *ligga*
refer to situations where a physical object rests on a supporting surface,
whereas *sitta* is typically used when such support is missing. The
attachment can be of various types such as adhesion (see 10) or a tight fit
that "holds" an object in its place (see (11)).

(10) Sedan ritade jag rent spårplanen för just den stationen och den
industrin och klistrade in fotografierna där de skulle *sitta*. (JMY1)
Then I drew track plans for each station or factory and pasted in
the photographs where they should *be*. (JMY1)

(11) Där *satt* redan de två knivarna i sina slidor. (HM1)
The two knives *were* already there in their sheaths. (HM1T)

The role played by the concept of attachment is very language-specific in
Swedish and is not lexicalized in the same way in closely related
languages such as German and Dutch (Viberg 2013: 160-163).
     Both the English and the Swedish postural verbs are frequently used
with **abstract** subjects (*Inflation stands at three percent*), but such uses
are based on conceptual metaphors, which are often lexicalized as more
or less fixed phrases. For that reason they follow different types of rules
and require separate treatment. To a certain extent uses with such
subjects are shared between English and Swedish, but very often
translations contain verbs from more abstract fields and the translation is
in many cases rather free, as in (12).

(12)   Unfortunately this theory does not *stand up* too well either. (DM1)
Tyvärr *fungerar* inte heller detta särskilt bra. (DM1T)
[Unfortunately, this does not function very well either]

Various types of subject are more or less marked or prototypical. Human subjects represent the unmarked or prototypical category. The use of postural verbs with Human subjects is conceptually basic and the use with other types of subject is derived via metonymy and/or metaphor.

3.2 More specific categories
It is possible to draw finer distinctions and to distinguish a number of further subcategories, but only a few more specific categories will be discussed. In many other cases, the examples are too few to make a quantitative comparison meaningful. The categories that will be identified are "robust" and can be coded and counted in a straightforward way. The category inanimate (concrete) subject can be divided into a number of subcategories. The basic contrasts represented in Table 3 apply most clearly to **physical objects** that can be held in the hand and manipulated, such as *bottle* or *plate*. Another important subcategory is **building**. According to Newman (2009, 46-52), *stand* and *sit* are used in English to refer to the location of buildings, and both of these alternatives occur in the ESPC corpus. In Swedish, *ligga* is the dominant alternative as in the translation of example (13). *Stå* can also be used, when the subject refers to a building, but that is a less frequent alternative.

(13)   The house itself was medium-sized, unexceptional to look at, *standing* on a street of such houses in an older part of Baltimore. (AT1)
Själva huset var medelstort och skilde sig inte från andra där det *låg* vid en gata med hus av samma typ i en äldre del av Baltimore. (AT1T)

This said, there are still some examples with *lie* in English originals as in (14).

(14)   Corte's farm *lies* just north of the Panama Canal watershed. (LT1)
Cortes gård *ligger* strax norr om Panamakanalens vattenavrin-ningsområde. (LT1T)

It must be pointed out that the most frequent correspondent to *ligga* in the ESPC corpus is locative *be* (see 15).

(15)   Hans hus *ligger* inte tvåhundra meter från hennes. (SW1)
        "His house *is*n't two hundred yards from hers. (SW1T)

In addition to straightforward examples meaning 'house', 'building', 'church' 'farm' and 'cottage' (Swed. *stuga*), the category building also includes parts of buildings such as apartments (see (16)) and rooms (e.g. 'kitchen') and nouns such as 'theatre', 'café', 'restaurant' and 'bakery' (see (17)), which may be parts of buildings or constitute free-standing buildings.

(16)   Lägenheten *låg* högst upp i hyreshuset. (KOB1)
        His apartment *was* on the very top floor of the block. (KOB1T)

(17)   Han erinrade sig att *det låg* ett konditori på gatan som gick parallellt med järnvägen. (HM2)
        He remembered *there was* a bakery on the street running parallel to the railroad. (HM2T)

A further category is **geographical place** (for short called place in the diagrams below) such as countries and cities and natural formations such as mountains and fields (see (18) and (19)).

(18)   Triana *ligger* på "andra" sidan av floden Guadalquivir. (BTC1)
        Triana *is situated* on the "wrong" side of the Guadalquivir River. (BTC1T)

(19)   Runt ån *låg* ett otillgängligt myrland genomkorsat av djurstigar.(KE1)
        All round the river *was* inaccessible marshland crisscrossed by animal paths. (KE1T)

In Swedish, *stå* is frequently used with reference to writing in a way that has no correspondence in English, which justifies distinguishing **writing** as a separate category (see Section 4.1). Originally, the justification for using *stå* appears to be the orientation of the letters, but in present-day

Swedish, reference can be made either to the actual text, as in (20), or to the propositional content, as in (21).

(20)   I slutet på mormors brev som följde med presenterna brukade det *stå*: "Morfar hälsar så mycket." (MG1)
       At the end of Grandmother's letters, which accompanied the presents, it usually *said*: "Grandfather sends his best." (MG1T)

(21)   Men det måste *stå* på skylten hur långt det är till fornminnet. (SC1)
       [it must stand]
       But it ought to *say* on the sign how far it is to the ancient monument. (SC1T)

Writing has a special semantic status combining concrete and abstract readings (see Viberg 2013: 157-159 for discussion), but for the purposes of this paper, it is enough to say that examples of this type are treated separately and not included in the abstract (or inanimate) category.

4. Over- and under-representation across different classes of subjects

4.1 STAND

Diagram 4a compares the frequencies of the major categories of Swedish *stå* in originals and translations. As can be observed, all categories except Abstract subject are significantly under-represented in the translations. (Animal is too small to be meaningfully discussed.) Note in particular the under-representation of the category Writing, which is relevant only for Swedish *stå*.

    The relationships between originals and translations in English are shown in Diagram 4b. Human and Inanimate subjects are over-represented. Writing, on the other hand, which is a category that is unique to Swedish (with respect to English), only occurs in one translated example (0.13 per 100 000 words), and this means that calques or loan translations are practically non-existent in this category. Human and Inanimate subjects, which have overlapping functions in English and Swedish, differ with respect to frequency of use in originals and translations. In Swedish, no translation effect was found for Abstract

subjects and as will be evident, Abstract subjects do not follow the general pattern of other verbs either (for some discussion, see Section 6).

	Human subject	Inanimate subject	Writing	Abstract subject
■ Original	78	34	14	15
■ Translation	57	22	10	15

p<	.0001	.0001	.03	ns
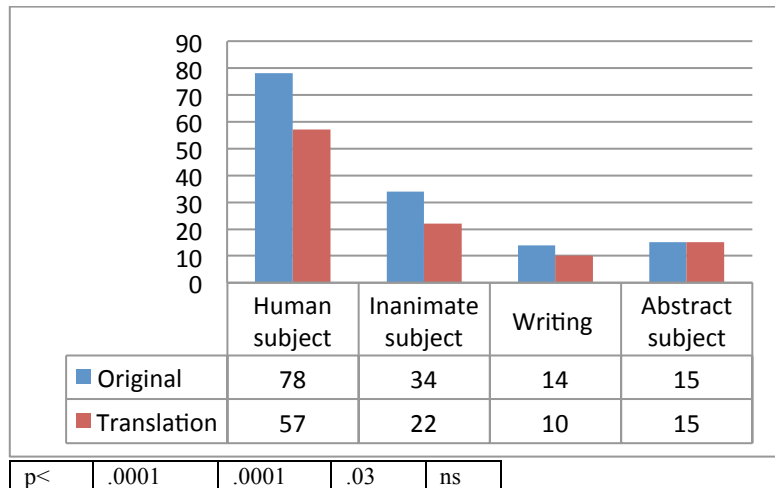
Diagram 4a. Major categories of Swedish Swedish *stå* in originals and translations. Frequency per 100,000 words.

	Human subject	Inanimate subject	Writing	Abstract subject
■ Original	32	6	0	3
■ Translation	50	11	0	3

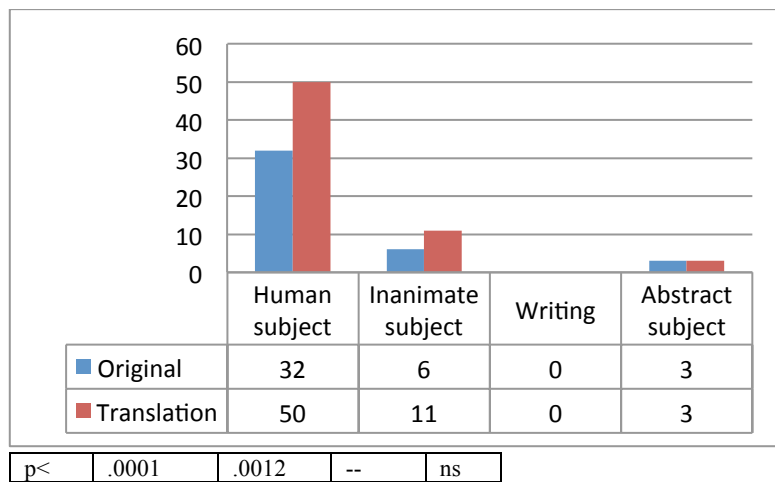| p<	.0001	.0012	--	ns

Diagram 4b. Major categories of English *stand* in orignals and translations. Frequency per 100 000 words.

By comparing Diagram 4a and 4b, we can also see that Human subjects and in particular Inanimate subjects are considerably more frequent in

Swedish than in English originals. In spite of that, both categories of use must be said to be clearly established in English. The overlap of functions appears to be a major explanation for the strong cross-linguistic influence on the translations.

4.2 LIE

Human subjects follow the same pattern for LIE as for STAND. They are under-represented in Swedish translated texts and over-represented in English translated texts. The same applies to Inanimate subjects taken as general category. However, in this section an attempt will be made to see what happens if different subcategories of Inanimate are distinguished. The reason for this is that *ligga* can be regarded as a default category for describing the location of inanimate subjects in Swedish. The frequency of occcurrence per 100 000 words for inanimate subjects of *ligga* is 56 in originals, which is higher than for human subjects of *ligga* and higher than for inanimate subjects of *stå* and *sitta*.

An inspection of the Swedish originals in Diagram 5a shows that physical object, building and place are well established in Swedish. In translations, Human is significantly under-represented. As for the subcategories of inanimate subjects, only Building and Other are significantly under-represented in the translated texts if tested individually. (Inanimate as a whole will be treated in 4.4.)

It is important to note that all the subcategories of Inanaimate distinguised in the analysis of Swedish also appear in the English original texts (see Diagram 5b), but all categories have a considerably lower frequency than in Swedish (compare Diagram 5a). There is substantial functional overlap between the systems, but salient differences with respect to frequency of use. In English translated texts, Human and the Inanimate subcategory Physical object are significantly over-represented, whereas Place and in particular Building (p just above 0.05) appear to be under-represented, but the data are inconclusive with respect to categories other than Physical object due to the low frequency in both originals and translations.

	Human	Physical object	Building	Place	Other inan.	Abstract
Original	29	19	18	10	9	26
Translation	21	16	14	11	5	22

p<	.001	.ns	.04	ns	.0009	ns
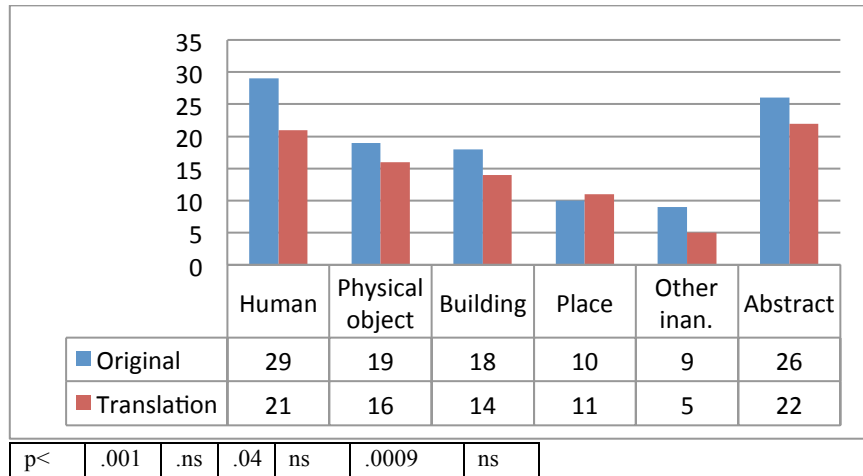
Diagram 5a. Categories of Swedish *ligga* in originals and translations. Frequency per 100 000 words.

	Human	Physical object	Building	Place	Other inan	Abstract
Original	10	4	2	4	2	7
Translation	20	8	1	3	3	7

p<	.0001	.0008	(.053)	ns	ns	ns
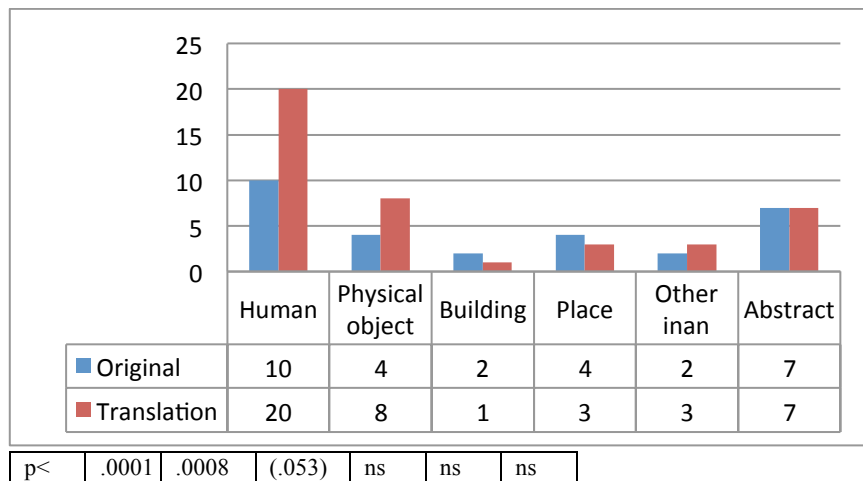
Diagram 5b. Categories of English *lie* in originals and translations. Frequency per 100 000 words.

4.3 SIT

The categories of Swedish *sitta* are presented in Diagram 6a. With respect to the use with inanimate subjects, *sitta* represents the most language-specific category among the postural verbs (see comments to Table 3 above). Although well established, the use of *sitta* with inanimate

subjects is marked even in comparison to the other postural verbs in Swedish, and is considerably less frequent than the use of *stå* and *ligga* with inanimate subjects. Per 100 000 words, its frequency of occurrence is 9, which should be compared to 34 for *stå* and 56 for *ligga*. Abstract subjects are infrequent with *sitta* in Swedish.

Diagram 6b shows the distribution of the categories of English *sit*. The most striking characteristic is that human subjects dominate completely over the other types. The other categories have such low frequencies that it is not meaningful to compare originals and translations. Human subjects are significantly over-represented in translations ($p<0.0001$).

	Human subject	Inanimate subject	Abstract subject	Animal subject
■ Original	72	9	1	2
■ Translation	61	5	0	2
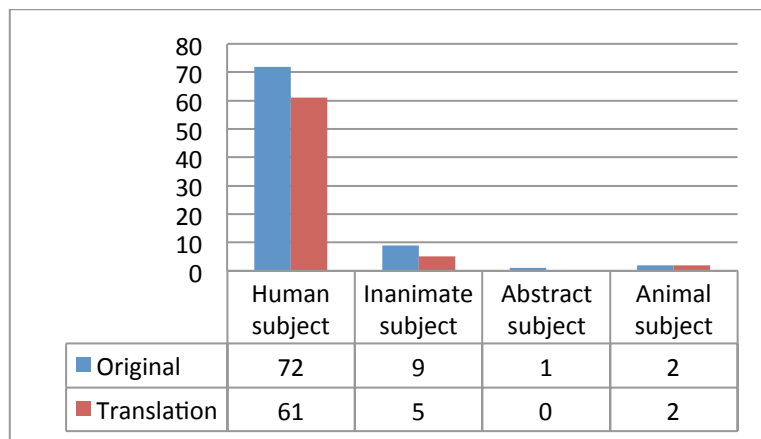
p<	.04	.009	na	ns

Diagram 6a. Categories of Swedish *sitta* (sit) in originals and translations. Frequency per 100 000 words

	Human subject	Inanimate subject	Abstract	Animal subject
■ Original	46	1	0	1
■ Translation	65	1	0	2
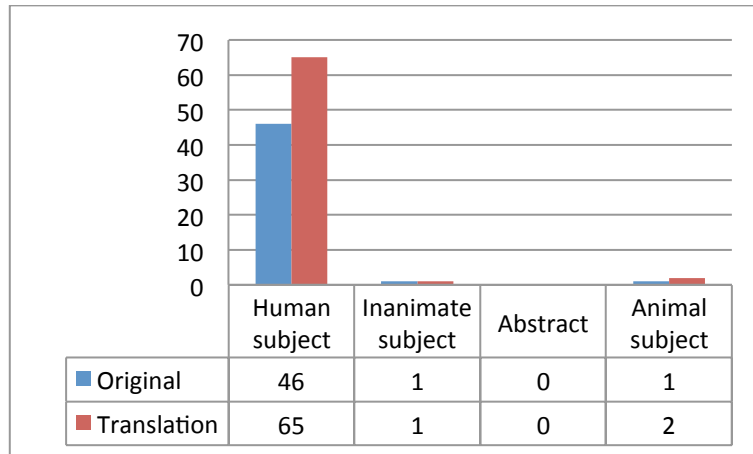
Diagram 6b. Categories of English *sit* in originals and translations. Frequency per 100 000 words

4.4 Contrast and convergence

Repeatedly, the clearest difference has been found between Human and Inanimate subjects. In order to illustrate this, the frequencies of occurrence of human and inanimate subjects are compared for all verbs in Diagram 7. In originals (Diagram 7a), there is contrast. Both human and inanimate uses are much more frequent in Swedish than in English, and this applies to all verbs. All of these differences are statistically significant except for SIT when it has an inanimate subject since it is too infrequent in English, but the difference is still clear. (The raw scores are 58 for *sitta* vs 4 for *sit*).

In translations (Diagram 7b), there is a significant statistical difference between the frequencies of human subjects only for STAND, and even in that case the difference is proportionally much smaller than for the inanimate subjects, which all are significantly more frequent in Swedish than in English.[4] There is a tendency for Human to converge in translations, whereas there is still a clear difference between the two languages in the Inanimate category.

---

[4] The reason why it is significant for SIT in translations but not in originals is that the raw scores are 37 vs 6 and that the test requires both scores to be equal to or higher than 5.

	Human STAND	Human LIE	Human SIT	Inan. STAND	Inan. LIE	Inan. SIT
Swedish	78	29	72	34	56	9
English	32	10	46	6	12	1
p<	.0001	.0001	.0001	.0001	.0001	na
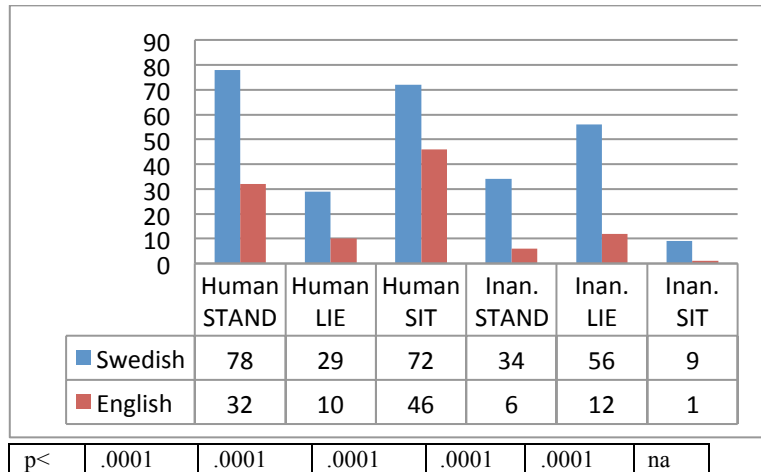
Diagram 7a Human and inanimate Subjects in Originals. Frequency per 100 000 words.

	Human STAND	Human LIE	Human SIT	Inan. STAND	Inan. LIE	Inan. SIT
Swedish	57	21	61	22	46	5
English	50	20	65	11	15	1
p<	.0008	ns	ns	.0001	.0001	.0001
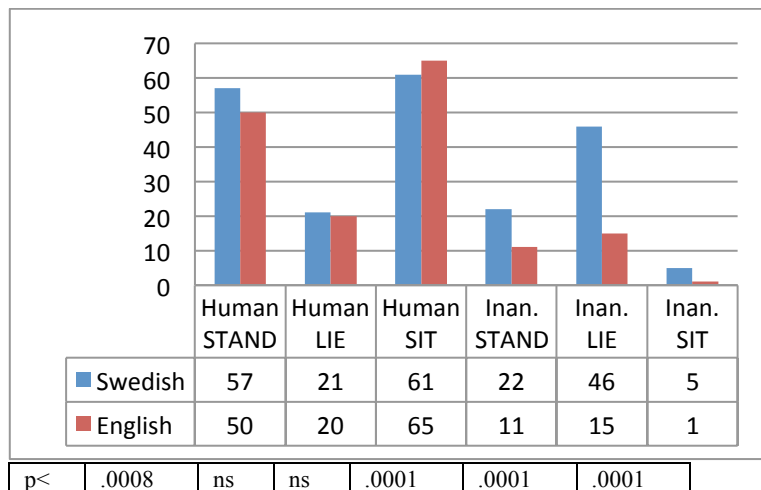
Diagram 7b. Human and inanimate subjects in translations. Frequency per 100,000 words

5. *What happens in translation: Translation shifts*

So far, the analysis has been restricted to the use of the corresponding verbs in originals and translations. What happens in translation, however, is much more complicated than that. This section will present a more qualitatively-oriented description of the translation pattern of individual

words using the categories discussed in Section 1.4. For reasons of space, only STAND will be systematically described. The general patterns are similar for LIE and SIT (see also Viberg 2013 for a contrastive study of the Swedish postural verbs).

5.1 Lexical shifts
The lexical shifts related to English *stand* when it is translated into Swedish are shown in Table 4. (The structural shifts will be commented on in Section 5.2.) As would be expected, direct lexical translations dominate.

Table 4. The translation pattern of English *stand* 1: Direct lexical translations and Lexical shifts:

| Type of subject | F | Translations | | | | |
| | | Direct | Lexical shifts | | | Other |
		(stå)	ställa sig	resa sig	ligga	
Human						
State/Activity						
Literal	151	133	4			14
Figurative	39	19	1			19
Change of Posture	40	1	13	19		7
Animal	6	5	1			0
Inanimate						
Physical object	29	20				9
Building	7	1			4	2
Other Literal	2	2				0
Figurative	10	4				6
Writing	0					
Abstract subject	17	5				12
Total	301	190	19	19	4	69

Since it turns out that direct lexical translations are much less frequent when the postural verbs are used figuratively (Table 4), literal

and figurative uses are accounted for separately when the subject is Human or Inanimate as in (22) and (23).

(22)   He was prepared to *stand by* that report. (FF1)
            Han var beredd att *hålla fast vid* rapporten. (FF1T)
            [hold on to the report]

(23)   One of the many tank-traps that *lie ahead*: senility's revenge. (JB1)
            En av de många fallgropar som *väntar*: senilitetens revansch.
            (JB1T)
            [that wait]

It should be noted that Figurative refers to the interpretation of the whole sentence. The subject in (22) still literally refers to a human. In (23), the subject *tank-trap* is used in a figurative sense, but the subject is classified as inanimate rather than as abstract since the choice of postural verb is still decided by the type of concrete inanimate referent that is referred to when tank-trap is used literally. Metaphorization applies to the whole proposition 'tank-traps lie ahead'.

    The major lexical shifts shown in Table 4 are *ställa sig upp* and *resa sig upp*, which refer to change-of-posture in Swedish. In English, change-of-posture is primarily formed by combining *sit*, *stand* and *lie* with a directional particle (e.g. *lie down*, *stand up*. See Newman 2009: 36-46 for discussion and some exceptions). One major difference between English and Swedish is that Swedish *sitta, stå* and *ligga* are not generally used to denote change of posture. Instead, a reflexive form of a related causative verb is used as in (24) or in some cases as in (25) a reflexive form of another action verb (*resa sig* 'rise', a reflexive of *resa* 'raise').

(24)   Kevin *stood up* again and scouted for a watchman. (RDO1)
            Kevin *ställde sig upp* igen och spanade efter vakten. (RDO1T)

(25)   Archbold *stood up* when Wexford came in. (RR1)
            Archbold *reste sig* när Wexford kom in, (RR1T)

The directional particles *upp* and *ner* 'down' are optional in Swedish with a few exceptions as indicated in Table 5, which shows the dynamic

system of Swedish posture verb (state/activity – postural change – causative).

Table 5. The dynamic system of Swedish Posture verbs (with Human subject).

State/Activity	Postural change (Inchoative)	Causative
stå 'stand'	*ställa sig* (upp) 'stand up' *resa sig* (*upp*) 'rise (up)'	*ställa* 'stand'
ligga 'lie'	lägga sig (ner) 'lie down'	lägga 'lay'
sitta 'sit'	*sätta sig* (ner)'sit down' *sätta sig upp* 'sit up' (from a lying position) *slå sig ner** 'sit down'; 'settle'	*sätta* 'seat', 'set'

*For *slå sig ner* 'sit down', literally 'strike oneself down', see Viberg (2016)

There are also a few examples where a Swedish postural verb is used in combination with a spatial particle to describe change of posture. This occurs in particular in the imperative form (*Stå upp*! (LH1) - *On your feet*! (LH1T)). [5]

Table 6 shows what happens when Swedish *stå* is translated into English. A direct translation is the most frequent alternative when the subject is human or inanimate and does not have a figurative meaning. Direct translations are used with a very low frequency in figurative uses with human subjects and with abstract subjects. Such translations require separate treatment and are only briefly discussed in this paper.

When the subject is human or animate, the second most frequent translation is *be*, used in various functions. This verb thus represents the most frequent type of lexical shift. The use of *be* to indicate the location of the subject is referred to as "Locative be" in Table 6. Proportionally, locative *be* occurs more often with inanimate than human subjects. The same applies even more to *there is* (in various inflected forms), which is treated separately and has been classified as a lexical shift. Sometimes it also involves a structural shift, when the Swedish verb does not appear in a presentative construction. (Although relevant, this is not shown in the table, since perspicuity would be lost if too many subcategories were to be introduced.)

---

[5] This use has such a low frequency (3 for *stå* in the originals) that it is not accounted for separately in Table 6.

In (26), the Swedish posture verb is used in a copula-like construction with an adjective or past participle (see 27) as a predicative complement. Such constructions occur both with human and inanimate subjects, which are coded separately but will be briefly commented on together in this section.

(26)   – Han *låg medvetslös* i över tjuge minuter, sa Åke. (KE1)
         [he lay unconscious]
         "He *was unconscious* for over twenty minutes," said Åke. (KE1T)

(27)   Dom *ligger begravda* vid kyrkan som alla andra. (SC1)
         [they lie buried]
         They*'re buried* beside the church like everybody else. (SC1T)

In most cases, English uses the copula as a translation ("Copula *be*" in Table 6), but English postural verbs can also appear in a copula-like construction. Example (28) is taken from an English original text.

(28)   The front door *stood open*.(RR1)
         Ytterdörren *stod öppen*. (RR1T)

Say is the most frequent translation of *stå* when the subject refers to Writing (see examples 20 and 21 above). Other relatively frequent alternatives specific to this category are *read* and *write* (usually in passive form: *be written*).

Table 6. The translation pattern of Swedish *stå*: Direct lexical translations and Lexical shifts

Type of subject	F	Lexical translations	Lexical shifts							Total lexical trans-lations
		Direct								
		stand	Locative be	Copula be	there is	say	read	write	Other	
Human										
Concrete	400	287	27	5	2	0	0	0	26	347
Figurative	117	8	16	7	0	0	0	0	33	64
Animal	4	2	0	0	0	0	0	0	0	2
Inanimate										
Concrete	209	52	39	17	26	0	0	0	26	160
Figurative	17	2	2	6	0	0	0	0	1	11
Writing	91	1	4	2	6	24	10	8	13	68
Abstract	101	11	4	14	1	0	0	0	31	61
TOTAL	939	363	92	51	35	24	10	8	130	713

5.2 Structural shifts

Structural shifts can have an impact on the occurrence of lexical correspondents in the translation. This section will discuss the types of structural shifts that directly affect the occurrence of postural verbs in the translations, especially various types of reductions that are characteristic of translations from Swedish into English. The clearest example is the reduction of relative clauses to prepositional phrases (or occasionally to some other type of place adverbial): *som* (Relative marker) + Postural verb + PP^{Locative} > PP^{Locative} (for short: Rel > PP). Relative clauses are typically introduced with the indeclinable relative marker *som,* as in (29).

(29)   En av orsakerna var hans far som bodde ensam i ett hus *som låg på slätten* strax utanför Löderup. (HM2)
        [that lay on the plain]
        One of the causes was his father, who lived alone in a house *on the plain* just outside Löderup. (HM2T)

There are 18 shifts of this type when *stå* is translated into English. This kind of reduction is mirrored by expansion when English is the source language as in (30). (There are 12 such expansions with *stå*.)

(30)   We passed a lady in her front garden. (JB1)
        Vi passerade en tant *som stod* i sin trädgård. (JB1T)
        [that stood in her garden]

Shifts of this type are particularly frequent with *ligga* when it has an inanimate subject (for *ligga* with inanimate subject, there are 31 reductions in English translated texts versus 34 expansions in Swedish translations). Example (31) goes from Swedish to English and (32) shows what happens in the opposite direction.

(31)   Jag pekade ut de få segelbåtarna *som fortfarande låg i vattnet.* (BL1)
        [that still lay in the water]
        I pointed out the few sailing-boats *still in the water*.(BL1T)

(32)   The memorandum *on his desk* when he arrived indicated this was not going to happen. (FF1)

Meddelandet *som låg på hans skrivbord* när han kom dit (…) (FF1T)
[that lay on his desk]

Actually, it is possible to find parallel examples where a postural verb without a correspondent in the Swedish version appears in a relative clause in English, as in (33), but there are very few of them. For *lie*, there is this example in the translated texts but there is none in the original English texts.

(33)   Vi måste gå till ett garage *litet längre bort*. (MS1)
        [a garage (a) little further away]
        We 've got to go to the garage *which lies a bit further back.* (MS1T)

The correspondents of *sitta* behave in a similar way as the other postural verbs (10 reductions of the type Rel > PP in translations from Swedish into English and 17 expansions PP > Rel in the opposite direction). Somewhat surprisingly, translational shifts involving relatives work in a direction that reduces the characteristic degree of over- and under-representation. In English translated texts, for example, postural verbs are over-represented in spite of the fact that reductions cut down the number of postural verb correspondents. The shifts involving relative clauses probably reflect general structural differences between English and Swedish, since relative clauses are very easy to form in Swedish and probably also are more frequent than in English.[6]

Another general structural difference affecting the translation patterns of the postural verbs are various types of participial constructions. Sometimes Swedish relative clauses are reduced to participial constructions in English, as in (34).

(34)   Trots att han försökte leva så anonymt som möjligt, i detta hus *som låg* strax öster om Ystad, hände det att nyfikna människor höll honom under uppsikt.(HM1)
        [this house that *lay* just east of Ystad]

---

[6] I am not aware of any systematic study of this, so this represents a suggestion for further studies.

> Even though he tried to live as anonymously as possible in this house *located* just east of Ystad, sometimes curiosity-seekers spied on him. (HM1T)

Such constructions are not completely ruled out in Swedish but usually sound very formal (*ett hus beläget strax öster om Ystad)*. There is another type of structural shift that involves a past participle, where a postural verb is translated with a resultative expression as in (35).

(35)   Vintaffeln *står* i Karl XIV Johans sängkammare. (GAPG1)
        [The wine table stands]
        The wine table *is placed* in the bedroom of Karl XIV Johan.
        (GAPG1T)

Some expressions included into this structural shift such as *be located* (see (36)) and *be situated,* which are used primarily as translations of *ligga* with places and buildings as subjects, are so conventionalized and have such a bleached resultative meaning that it would be justified to regard them as purely lexical (multi-word) shifts.

(36)   Den hette Ercildoune och *låg* utanför Centerport (JMY1)
        [lay outside Centerport]
        It was called Ercildoune and *was located* outside Centerport
        (JMY1T)

Non-finite *ing*-clauses and *ed*-clauses in various syntactic roles (Biber et al 1999, 199-200) can very often not be translated with the corresponding present and past participles in Swedish without sounding unidiomatic. This lack of correspondence relatively frequently leads to reductions and expansions of various types. In (37) and (38) the postural verb disappears in the Swedish translation.

(37)   *Lying* just south of the Thames in west London, Richmond Park is
        the most "natural" and largest of the London Royal Parks (SUG1)
        Strax söder om Themsen i sydvästra London, (…) (SUG1T)
        [Just south of the Thames…]

(38)   Jim Rawlings spent the hour between nine and ten that night

sitting in another, smaller rented car outside Fontenoy House. (FF1)
Jim Rawlings tillbringade timmen mellan nio och tio den kvällen [*Zero*] i en annan, mindre hyrbil utanför Fontenoy House. (FF1T)
[spent the hour --- in another --- car]

Another option is to expand the non-finite clause into a relative clause (if the structural context allows it), as is done with the *ed*-clause in (39) in a way that introduces a postural verb into the translation.

(39)   (…) music so loud it made the transistor radio *balanced* on the stoop steps buzz and tremble. (NG1)
         (…) transistorradion *som stod balanserad* på förstubron (NG1T)
         [that stood balanced]

The structural shifts characteristic of the translations of non-finite clauses are a general phenomenon that affects the grammar more than the lexicon, but still the effect on the translation patterns of postural verbs appears to be rather great.[7]

5.3 Swedish pseudo-coordination

The most discussed characteristic of Swedish postural verbs is their use in so-called pseudo-coordination (see Darnell 2008 for an overview of earlier research and a detailed study). This structure can be regarded as a case of emergent grammaticalization and has often been compared to the general tendency across languages to develop durative aspectual meanings such as the progressive and the continuative (Newman 2002). Recently, Blensenius (2015) has challenged the view that pseudo-coordination with postural verbs expresses progressive meaning in Swedish. They are "instead locative and, in some cases, episodic" (Blensenius 2015, Abstract).

In pseudo-coordinations, the postural verb is de-accented and coordinated with a lexical verb, as in (40). There are at least four recurring types of English correspondents of pseudo-coordination. The correspondences go in both directions, which will be demonstrated by

---

[7] Such shifts have been coded in the database but need to be worked on more in order to present reliable statistics.

presenting first an example from a Swedish original text and then from an English original. The most frequent and also the closest correspondent, according to my intuition, which is not as strong for English as it is for Swedish, is the construction $V^{Postural}$ + V-ing, which appears in (40). The construction is referred to as "simultaneous conjunction" and is regarded as "a potential root of grammaticalization" in Newman & Rice (2004, 371).

(40)   När jag *låg och smälte frukosten* uppenbarade sig ronden. (PCJ1)
       [lay and melted the breakfast]
       While I *lay digesting my breakfast*, the round appeared. (PCJ1T)

The same type of correspondence can be observed when English is the source language (see 41).

(41)   He would drop the spoon in the sink and *stand sipping* from his mug while the cat wove between his feet. (AT1)
       Sedan skulle han släppa ner skeden i diskhon och *stå och läppja* på kaffet medan katten slingrade sig ut och in mellan fötterna på honom. (AT1T)

However, there are several other relatively frequent correspondences. One option is to use a single lexical verb in the progressive, as in (42) and (43). This happens 21 times when *stå* is translated and 17 times when *sitta* is translated (from Swedish originals) but only twice when *ligga* is pseudo-coordinated.

(42)   Patron Björk kommer in en dag när han *stod och rätade till* flugan: (…) (GT1)
       [when he stood and straightened his bow tie]
       Boss Björk came in one day when he *was straightening* his bow tie: (…) (GT1T)

(43)   He was replacing the element in the kettle and she *was cutting up* vegetables for one of her delicious cheap dishes; (NG1)
       Han lagade elementet i tekitteln och hon *stod och skar* grönsaker till delikata billiga rätter. (NG1T)
       [stood and cut]

Further options are to use a co-ordination also in English, as in (44) and (45).

(44)   Jag gick ut i väntrummet där min far *satt och väntade*. (MS1)
        I went into the waiting room where my father *sat and waited.*
        (MS1T)

(45)   I *sat and stared* at the flies. (BO1)
        Jag *satt och stirrade* på flugorna. (BO1T)

There are also examples where a single lexical verb in non-progressive form appears in the English version as in (46), but this option has a rather low frequency.

(46)   As he was borne upwards in a series of disconcerting jerks
        Dalgliesh *reflected* that success, (…) (PDJ1)
        Medan hissen ryckvis förde Dalgliesh uppåt *stod* han *och tänkte på*
        att framgången, (…) (PDJ1T)

Pseudo-coordination has been commented on rather extensively because of its theoretical interest, but it will not be studied in detail in this paper. In order to be meaningful, a study would require a detailed discussion of the problem to identify all pseudo-coordinations in an actual corpus, which appears to have been underestimated.[8] In the quantitative tables, postural verbs appearing in pseudo-coordinations are simply counted like other postural verbs, since they still refer to posture in this construction.

6. The problem with genre, in particular Non-Fiction
The comparisons between originals and translations in Section 4 are based on the whole corpus. As mentioned in Section 2.3, postural verbs have an overall frequency that is much higher in Fiction than in Non-Fiction. This section will show what happens when Fiction and Non-Fiction are analyzed separately. Only raw frequencies will be presented,

---

[8] That is why no quantitative table is presented in Section 5.3. What is said about frequencies is based on a coding of all examples that tentatively have been identified as pseudo-coordinations.

since the frequencies in many cases would be extremely low if presented per 100 000 words. (As in Section 4, the statistical testing is based on the raw frequencies.)

6.1 The distribution of meanings across genres
The differences between genres are not restricted to the overall frequency, but also involve the distribution of meanings or uses. As can be observed in Table 7, Non-Fiction contains a greater proportion than Fiction of human subjects in sentences with a figurative meaning such as *I can't stand him* and an even greater proportion of abstract subjects (except for *sitta*, which is primarily used literally). Direct translations dominate when the subject is human and the overall interpretation is literal, reaching a proportion as high as 71% if the result is summed up for all three postural verbs. For figurative uses of Humans subjects and for Abstract subjects, the proportion of direct translations is very low.

Table 7. The distribution of meanings across genres in Swedish original texts

Type of Subject	stå Fiction		Non-Fiction		ligga Fiction		Non-Fiction		sitta Fiction		Non-Fiction	
	Total	Dir	Total	Dir	Total	Dir	Total	Dir	Total	Dir	Total	Dir
Human												
Literal	357	265	43	18	149	103	10	5	373	276	95	63
Figurative	39	8	78	5	22	4	8	0	9	3	7	2
Animal	4	2	0	0	7	5	0	0	10	7	0	0
Inanimate												
Literal	175	48	34	5	264	71	82	6	48	5	10	0
Figurative	7	1	10	1	12	1	7	1	0	0	0	0
Writing	61	0	30	1	--	--	--	--	--	--	--	--
Abstract	26	2	75	9	33	4	139	20	2	0	3	0
Total	669	326	270	39	487	188	246	32	442	291	115	65
% Direct		49		15		39		13		66		58

6.2 The influence of genre on over- and under-representation

In this section, the comparisons will be made separately for Fiction and Non-Fiction. Table 8 shows the result for STAND. Fiction (see Table 8ab) follows the pattern of the complete corpus, whereas the statistical test in many cases is not applicable or gives a non-significant result when we look at Non-Fiction (see Table 8cd). In addition, there is at least one significant result in Non-Fiction that contradicts the general pattern, since *stand* is significantly under-represented in English translations when the subject is Human. One explanation for this could be that figurative uses of Human are much more frequent in Non-Fiction than in Fiction, and, as demonstrated above, figurative uses behave in a different way than literal uses. In particular, there are few direct translations.

Table 8a English *stand* in originals and translations, fiction

	Number of occurrences		Significance of difference
	Original	Translation	
Human	177	341	<0.0001
Animal	3	3	n.a.
Inanimate	35	68	<0.0004
Writing	0	0	n.a.
Abstract	5	7	n.s.
Total	220	419	<0.0001
Corpus size	340,745	333,375	

Table 8b Swedish *stå* in originals and translations, fiction

	Number of occurrences		Significance of difference
	Original	Translation	
Human	396	275	<.0001
Animal	4	3	n.a.
Inanimate	182	111	<.0001
Writing	61	54	n.s.
Abstract	26	44	<0.0482
Total	669	487	<.00001
Corpus size	308,160	346,649	

Table 8c English *stand* in originals and translations, non-fiction

	Number of occurrences		Significance of difference
	Original	Translation	
Human	53	34	0.0043
Animal	3	0	n.a.
Inanimate	13	16	n.s
Writing	0	1	n.a.
Abstract	12	16	n.s.
Total	81	67	0.0275
Corpus size	364,648	413,500	

Table 8d Swedish *stå* in originals and translations, non-fiction

	Number of occurrences		Significance of difference
	Original	Translation	
Human	121	126	n.s
Animal	0	4	n.a.
Inanimate	44	36	n.s
Writing	30	16	<0.0241
Abstract	75	61	n.s
Total	270	243	n.s
Corpus size	353,303	344,131	

Table 9 shows the result for LIE. In this case, Non-Fiction also presents some problems for the generalizations arrived at in Section 4. In Swedish, human subjects are significantly over-represented in Non-Fiction (see Table 9d). However, it must be noted that the number of examples is small. The proportion of figurative uses is also high (8 out of 18 in originals and 11 out of 30 in translations). Inanimate subjects also behave in an exceptional way and are significantly under-represented in English Non-Fiction and close to significantly over-represented in Swedish Non-Fiction. A closer look at the data reveals that the number of physical objects is small and the major difference is related to the category of subjects coded as Place (see Table 9e).

Table 9a English *lie* in originals and translations, fiction

	Number of occurrences		Significance of difference
	Original	Translation	
Human	57	139	0.0001
Animal	4	8	n.a.
Inanimate	37	91	0.001
Abstract	19	12	n.s.
Total	117	250	0.0001
Corpus size	340,745	333,375	

Table 9b Swedish *ligga* in originals and translations, fiction

	Number of occurrences		Significance of difference
	Orig.	Trans.	
Human	171	117	0.0001
Animal	7	11	n.s.
Inanimate	276	213	0.0001
Abstract	33	42	n.s
Total	487	383	0.0001
Corpus size	308,160	346,649	

Table 9c English *lie* in originals and translations, non-fiction

	Number of occurrences		Significance of difference
	Original	Translation	
Human	12	9	n.s
Animal	3	0	n.a.
Inanimate	39	12	0.0001
Abstract	33	38	n.s.
Total	87	59	0.001
Corpus size	364,648	413,500	

Table 9d Swedish *ligga* in originals and translations, non-fiction

	Number of occurrences		Significance of difference
	Original	Translation	
Human	18	30	0.0342
Animal	0	0	n.a.
Inanimate	89	109	(0.054)
Abstract	139	112	(0.0533)
Total	246	251	n.s
Corpus size	353,303	344,131	

Table 9e The major subcategories of inanimate for English *lie* and Swedish *ligga*, non-fiction.

	English *lie*		Swedish *ligga*	
	Original	Translation	Original	Translation
Physical object	5	5	15	9
Place	21	4	20	48
Building	9	1	44	41

SIT as shown in Table 10 is characterized by a dominance of human subjects, which are used literally (as shown in Table 7). Apart from that, both Fiction and Non-Fiction follow the pattern for the whole corpus and do not present any problems for the generalizations presented in Section 4.

Table 10a English *sit* in originals and translations, fiction

	Number of occurrences		Significance
	Original	Translation	
Human	276	401	<0.0001
Animal	10	12	n.s.
Inanimate	3	6	n.a.
Abstract	0	0	n.a.
Total	289	419	<0.0001
Corpus size	340,745	333,375	

Table 10b Swedish *sitta* in originals and translations, fiction

	Number of occurrences		Significance
	Original	Translation	
Human	382	372	<0.0237
Animal	10	9	n.s.
Inanimate	48	33	<0.0139
Abstract	2	0	n.a.
Total	442	414	<0.0036
Corpus size	308,160	346,649	

Table 10c English *sit* in originals and translations, non-fiction

	Number of occurrences		Significance
	Original	Translation	
Human	49	85	<0.0085
Animal	0	0	n.a.
Inanim.	1	0	n.a.
Abstract	0	0	n.a.
Total	50	85	<0.0111
Corpus size	364,648	413,500	

Table 10d Swedish *sitta* in originals and translations, non-fiction

	Number of occurrences		Significance
	Original	Translation	
Human	102	52	<0.0001
Animal	0	4	n.a.
Inanimate	10	4	n.a.
Abstract	3	2	n.a.
Total	115	62	<0.0001
Corpus size	353,303	344,131	

To sum up, the comparison across genres shows that the generalizations arrived at in Section 4 very clearly hold up for Fiction. For Non-Fiction, the picture is unclear. One reason for that could be the low proportion of literal uses. Another reason is that Non-Fiction is negatively defined. It actually contains a number of different types of texts such as biographies, academic prose, written versions of public speeches, legal texts and annual reports. For lexical studies, this variety is valuable for giving examples of a wider range of uses than what would have been possible with a more narrowly defined selection (given the size of the corpus). On the other hand, this variety makes it more difficult to achieve comparability across languages and makes quantitative comparisons as the ones presented here less reliable.

7. Conclusions and discussion
The major result of the present study is that there is a rather strong tendency in translations towards convergence in the usage patterns (word frequencies) between languages on points where functions/meanings overlap. Unmarked categories (Human subjects) tend to be more affected

than marked categories (Inanimate subjects). In spite of this, the system of semantic contrasts tends to be preserved.

As a test of the methodology used in corpus-based contrastive studies, the study shows that special care must be taken when mono-directional translation corpora are used and the characterization of some language(s) is based on translated texts alone (cf. Johansson 2007: 31-33). Word frequencies may be strongly affected in particular when the functions overlap across languages. On the other hand, the system of semantic contrasts is preserved, judging from the present study, which probably represents a particularly problematic case. The preservation of the semantic contrasts and the avoidance of translations that can directly be spotted as semantic calques presuppose that translations of high quality are used. Problems do not arise when bi-directional corpora such as the ESPC are used and translation patterns are used only as a complement to the comparison of original texts in both languages. From this, it appears that the inclusion of original texts from all the languages compared is ideal. However, translations make it possible to compare how a specific meaning is expressed in exactly the same context, so the exclusive use of original texts may also be problematic due to lack of comparability. When many languages are compared, as is the case in typologically oriented studies based on corpora, the desirability of having original texts in all languages must be balanced against the desirability of having access to translations, since resources in terms of time and effort are limited.

As a contribution to general linguistic theory, the result of the study of translation effects should be related to theories of bilingualism and language contact. Of special interest from the point of view of contact linguistics and language change is the comparison by Kranich, Becher & Höder (2011) of the contact between Latin and Old Swedish and the contact between English and German today and their discussion of the importance of factors such as the degree of standardization overall of the target language and the establishment of the genre in the target language. Old Swedish lacked a written standard. The influence of Latin could be traced at all linguistic levels, and Latin served as a model for many new genres to the extent that formal equivalence became a preferred strategy in translation at the end of the period (for example, in religious literature). In German, where two relatively new genres were studied (popular

science and business communication), the influence was restricted to pragmatic and stylistic features.

However, the convergence seen in translations (especially) of fiction in the present study is related neither to the lack of standardization of the languages involved nor to the lack of an establishment of the genre. In most cases, it is probably completely unconscious as far as the translators are concerned, and it does not fill any expressive gaps; nor is it likely to affect the structure of original texts for the time being. On the other hand, it is a latent feature probably present in most situations where translations are produced, and when the external conditions are right, this type of translation effect can lead to contact-induced language change. (This type of change is discussed in Thomason 2001, Chapter 4).

It is also interesting to compare translation effects with data from bilingual speakers and L2 learners. At a general level (with important differences at a more detailed level), the convergence in translations by professional translators noted in Section 4 has a parallel in second-language acquisition and in bilingual development. In a recent study of Turkish-German contact varieties, Goschler (2013, 127) concludes: "Most of the time, bilinguals neither simply transfer constructions from their L1 to their L2, nor do they behave exactly like monolinguals. Instead, they use L2 patterns that occur in their L1 as well". (Cf. bilingual "compromise systems", Obler 1982.)

Professional translators are bilingual speakers with a very advanced level of competence in both their languages. Ideally they translate into their strongest language. When comparing with L2 learners, it is important to take the level of proficiency into consideration, since acquisition often involves several different steps (see Viberg 2002 for a summary of the development of the causative counterparts of the postural verbs in Swedish used as placement verbs, *sätta/ställa/lägga* 'cause-sit/stand/lie'). In a longitudinal study of Swedish as L2, Viberg (1999) found that postural verbs belonged to the type of frequent verbs that had a tendency to be under-represented in the speech of the L2 learners in comparison to native controls. This was evident even in recordings carried out two years after the first recordings. (After four years, *sitta* was still under-represented, whereas the situation was unclear for *stå* and *ligga*.) More recently, Lemmens & Perez (2012) have presented a study of French learners of Dutch as L2 based on oral picture descriptions. Learners were divided into three proficiency groups and compared to a

group of native speakers. In general, the learners had a clear tendency to under-use Dutch postural verbs. Given that both Swedish and Dutch use postural verbs much more than French, this is a clear parallel to the under-use of Swedish postural verbs that Kortteinen (2008) found in translations from French (see Section 1.2). However, the French L2 learners also had a tendency to confuse the different posture verbs. Such confusion lacks a clear parallel in the translation data. There was also a tendency among learners at a more advanced stage to over-use some of the Dutch verbs. In general, there was a U-shaped development, since confusions increased in frequency as the learners started to use posture verbs more often.

Thus, what happens in translation sheds interesting light on what happens in language contact and in bilingual development (cf. Hyltenstam & Viberg 1993). If professional translators are regarded as highly proficient bilinguals, translation effects represent cross-linguistic influence at the upper end of bilingual development. In a similar vein, language contact can be regarded as the study of bilingual language development at the societal (in distinction to the individual) level. From this perspective, the translation effects found in this study represent a kind of language variation that is limited to translated texts, but that − if the social circumstances are the right ones − can rise to the surface and be established as part of the norm of the language.

References

Altenberg, Bengt and Karin Aijmer. 2000. "The English-Swedish Parallel Corpus: A Resource for Contrastive Research and Translation Studies." *Corpus Linguistics and Linguistic Theory*. Eds. Christian Mair and Marianne Hundt Amsterdam – Atlanta/GA: Rodopi. 15-33.

Baker, Mona. 1993. "Corpus Linguistics and Translation Studies – Implications and Applications." *Text and Technology. In Honour of John Sinclair*. Eds. Mona Baker, Gill Francis and Elena Tognini-Bonelli. Amsterdam: Benjamins. 233-250.

Becher, Viktor, Juliane House and Svenja Kranich. 2009. "Convergence and Divergence of Communicative Norms through Language Contact in Translation." *Convergence and Divergence in Language Contact Situations*. Eds. Kurt Braunmüller and Juliane Hause. Amsterdam: Benjamins.125-151.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Blensenius, Kristian. 2015. *Progressive Constructions in Swedish*. [Göteborgsstudier i nordisk språkvetenskap 25] PhD thesis. Department of Swedish, Gothenburg University.

Braunmüller, Kurt and Juliane Hause. Eds. 2009. *Convergence and Divergence in Language Contact Situations*. Amsterdam: Benjamins.

Darnell, Ulrika Kvist. 2008. *Pseudosamordningar i svenska: särskilt sådana med verben* sitta, ligga *och* stå. PhD thesis. Department of Linguistics, Stockholm University.

Ebeling, Jarle and Signe Okseefjell Ebeling. 2013. *Patterns in Contrast*. Amsterdam: Benjamins.

Englund Dimitrova, Birgitta. 2010. "Translation Process." *Handbook of Translation Studies*. *Vol. 1*. Eds. Yves Gambier and Luc van Doorslaer. Amsterdam: Benjamins. 406-411.

Gambier, Yves. 2010. "Translation Strategies and Tactics." *Handbook of Translation Studies*. *Vol. 1*. Eds. Yves Gambier and Luc van Doorslaer. Amsterdam: Benjamins. 412-418.

Gellerstam, Martin. 1986. "Translationese in Swedish Novels Translated from English." *Translation Studies in Scandinavia*. [Lund Studies in English 75] Eds. Lars Wollin and Hans Lindquist. Lund: CWK Gleerup. 88-95.

Gibbs, Raymond W., Dinara A. Beitel, Michael Harrington and Paul E. Sanders. 1994. "Taking a Stand on the Meaning of *stand*: Bodily Experience as Motivation for Polysemy." *Journal of Semantics* 11: 231-251.

Goschler, Juliana. 2013. "Motion Events in Turkish-German Contact Varieties." *Variation and Change in the Encoding of Motion Events*. Eds. Juliana Goschler and Anatol Stefanowitsch. Amsterdam: Bejamins. 115-131.

Halverson, Sandra. 2003. "The Cognitive Basis of Translation Universals." *Target* 15-2: 197-241.

Halverson, Sandra. 2010. "Cognitive Translation Studies. Developments in Theory and Method." *Translation and Cognition*. Eds. Gregory Shreve and Erik Angelone. Philadelphia: John Benjamins. 349-369.

Hyltenstam, Kenneth and Åke Viberg. Eds.1993. *Progression and Regression in Language*. Cambridge: Cambridge University Press.

Jarvis, Scott and Aneta Pavlenko. 2008. *Crosslinguistic Influence in Language and Cognition*. London and New York: Routledge.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora*. Amsterdam: Benjamins.

Kortteinen, Pauli. 2008. *Les verbes de position suédois* stå, sitta, ligga *et leurs équivalents français*. *Étude contrastive*. [Acta Univeristatis Gothoburgensis LX.] Göteborg: Göteborgs universitet.

Kranisch, Svenja, Viktor Becher and Steffen Höder. 2011. "A Tentative Typology of Translation-Induced Language Change." *Multilingual Discourse Production. Diachronic and Synchronic Perspectives*. Eds. Svenja Kranisch, Viktor Becher, Steffen Höder and Juliane House. Amsterdam: Benjamins. 11-43.

Laviosa, Sara. 2002. *Corpus-Based Translation Studies*. Amsterdam - New York: Rodopi.

Lemmens, Maarten. 2002. "The Semantic Network of Dutch Posture Verbs." *The Linguistics of Sitting, Standing, and Lying*. Ed. John Newman. Amsterdam: Benjamins. 103-139.

Lemmens, Maarten and Julien Perrez. 2010. "On the Use of Posture Verbs by French-Speaking Learners of Dutch: A Corpus-Based study." *Cognitive Linguistics* 21-2: 315-347.

Lemmens, Maarten and Julien Perrez. 2012. "A quantitative Analysis of the Use of Posture Verbs by French-Speaking Learners of Dutch." *CogniTextes*. Vol. 8. http://cognitextes.revues.org/609. [07 June 2013]

Malmkjær, Kirsten. 2011. "Translation Universals." *The Oxford Handbook of Translation Studies*. Oxford: Oxford University Press. Eds. Kirsten Malmkjær and Kevin Windle. 83-93.

Malmkjær, Kirsten and Kevin Windle. Eds. 2011. *The Oxford Handbook of Translation Studies*. Oxford: Oxford University Press.

Mauranen, Anna and Pekka Kujamäki. Eds. 2004. *Translation Universals*. Amsterdam: Benjamins.

Munday, Jeremy. 2008. *Introducing Translation Studies: Theories and Applications*. 2nd ed. London: Routledge.

Newman, John. Ed. 2002. *The Linguistics of Sitting, Standing, and Lying*. Amsterdam: Benjamins.

Newman, John. 2002. "A Cross-Linguistic Overview of the Posture Verbs 'sit'. 'stand' and 'lie'." *The Linguistics of Sitting, Standing, and Lying*. Ed. John Newman. Amsterdam: Benjamins. 1-24.

Newman, John and Sally Rice. 2004. "Patterns of Usage for English SIT, STAND, AND LIE: A Cognitively Inspired Exploration in Corpus Linguistics." *Cognitive Linguistics 15-3*: 351-396.

Newman, John. 2009. "English Posture Verbs. An Experientially Grounded Approach." *Annual Review of Cognitive Linguistics*: 30-57.

Obler, Loraine K. 1982. "The Parsimonious Bilingual." *Exceptional Language and Linguistics*. Eds. Loraine Obler and Lise Menn. New York: Academic Press. 339-346.

Stubbs, Michael and Isabel Barth. 2003. "Using Recurrent Phrases as Text-Type Discriminators. A Quantitative Method and some Findings." *Functions of Language* 10-1: 61-104.

Svensson, Kristina. 2005. *Uno studio contrastivo svedese-italiano sui verbi svedesi* stå, sitta *e* ligga: Thèse pour le doctorat. Göteborg: Göteborgs universitet.

Thomason, Sarah G. 2001. *Language Contact. An Introduction*. Edinburgh: Edinburgh University Press.

Viberg, Åke. 1998. "Crosslinguistic Perspectives on Lexical Acquisition: The Case of Language-Specific Differentiation." *Perspectives on Lexical Acquisition in a Second Language*. [Travux de l'institut de linguistique de Lund 38.] Eds. Kirsten Haastrup and Åke Viberg. Lund: Lund University Press. 175-208.

Viberg, Åke. 1999. "Lexical Development in a Second Language." *Bilingualism and Migration*. Eds. Guus Extra and Ludo Verhoeven. Berlin: Mouton de Gruyter. 165-185.

Viberg, Åke. 2002. "Basic Verbs in Second Language Acquisition." *Revue Française de Linguistique Appliquée* VII-2: 51-69.

Viberg, Åke. 2013. "Posture Verbs. A Multilingual Contrastive Study." *Languages in Contrast* 13-2: 139-169.

Viberg, Åke. 2016. "Polysemy in Action: The Swedish Verb *slå* 'hit, strike, beat' in a Crosslinguistic Perspective." *The Lexical Typology of Semantic Shifts*. Eds. Maria Koptjevskaja Tamm and Päivi Juvonen. Berlin: Mouton de Gruyter. 177-222.

Electronic sources

The English Swedish Parallel Corpus (ESPC). For a description, see: http://www.sol.lu.se/engelska/corpus/corpus/espc.html

Vassarstats: http://www.vassarstats.net

The Swedish modal particle *nog*. A contrastive analysis.

Karin Aijmer, University of Gothenburg

Abstract
The article investigates the meaning and functions of the Swedish modal particle *nog* on the basis of its cross-linguistic correspondences in the English-Swedish Parallel Corpus (ESPC). *Nog* was more frequent in original texts than in sources reflecting the fact that modal particles are used abundantly in Swedish conversations. *Nog* does not have an obvious correspondence in English as indicated by the fact that it has been rendered by many different lexical items and constructions. Moreover omission is a frequent strategy. The translations show that *nog* has two different core uses depending on position and stress. In medial position *nog* means both probability and (almost) certainty. When *nog* corresponds to an expression of certainty it can come to imply an element of self-assurance. Medial *nog* is also used to modify mental verbs (*jag vet nog* 'I know nog') especially in response utterances. An important function of the uses of *nog* in medial position is downtoning an opinion or an utterance which might be offensive to the hearer unless mitigated. *Nog* is also used deontically with reference to what ought to be done. The deontic meaning is particularly clear when *nog* is used with a modal auxiliary to give advice. *Nog,* when initial, involves emphasis or contradictory assumptions (contrastive *nog*). Contrasting opinions account for example for the translations of a sentence with initial *nog* by a negative interrogative sentence (or a declarative sentence with a tag question). Another function of the initial *nog* is to prepare the hearer for an objection in a following *but*-clause.

1. Introduction

Modal particles are found only in a few languages. They are notoriously difficult to analyse because of their multifunctionality and context-boundness. The angle chosen in the present study is a contrastive approach to the study of modal particles. The aim is to study the meaning and functions of the Swedish modal particle *nog* on the basis of its cross-linguistic correspondences in the English-Swedish Parallel Corpus (ESPC). Translations have probably always been used informally by the linguist to establish what a lexical element means in a context. In the present study this approach is generalized to large amounts of translation texts in two languages. We can therefore test hypotheses about what these meanings are based on the linguist's own intuitions or bilingual corpus data. The translations provide a 'rich' picture of the variability of

the modal particles which can be the basis for describing their polysemy and multifunctionality.

The paper is structured as follows. Section 2 defines modal particles (in general) and describes how they differ from modal adverbs and discourse markers. Section 3 deals with previous work on *nog*, and Section 4 describes the design of the English-Swedish Parallel Corpus and the contrastive method. Section 5 presents some quantitative results of the contrastive analysis. Section 6 discusses the functions of *nog* and how we should describe the lexical meaning of the particle. Section 7 provides a summary and conclusion.

2. Defining modal particles

Modal particles are abundant in German (see e.g. Diewald 2013, Waltereit 2001) and are also found in Swedish (Aijmer 1996, Aijmer 2015), Danish (Davidsen-Nielsen 1996), and Norwegian (Fretheim 1981, Borthen and Knudsen 2014).

As a group of words they have certain formal and functional characteristics distinguishing them both from modal adverbs and from discourse particles (see e.g. Diewald 2006, 2013, Waltereit 2001). Let us consider what (some of) these features are:

Formal features
Morphologically modal particles are 'particles'; they are 'non-inflected monosyllabic units that have segmental status and can be isolated as such' (Diewald 2007: 409). This distinguishes them from modal adverbs such as *probably*. Modal particles (unlike modal adverbs) are integrated in the sentence and have a fixed position in the so-called middle position after the finite verb.

The prototypical position of *nog* after the finite verb is illustrated in example (1):

(1)   Men det var <u>nog</u> bara prat.
      'But that was 'nog' only talk

Nog can also be found in initial position. However, it is still integrated in the utterance and it has a different function from the medial *nog:*

(2)   Nog är jag starkare än du
       'Nog' am I stronger than you.

Prosodic features
Modal particles are generally unstressed unlike modal adverbs (e.g. Diewald 2013). *Nog* presents special problems since it can also be sentence-initial and stressed.

Functional features
Generally speaking, modal particles express pragmatic meaning related to the attitude of the speaker and the hearer (Cuenca 2013: 195). However, depending on the context and their lexical meaning they have extended their meanings in different directions.

Stylistic features
Modal particles are found in spoken language rather than in writing which suggests that they have interactional functions in speech (Lindström 2008: 96).

Modal particles are difficult to distinguish from discourse markers. Both modal particles and discourse markers are for example used by the speaker to take up different positionings or stances in the interaction. A topical area in linguistics over the years has therefore been the intersection between modal particles and discourse markers and whether it is possible to draw a line between the two types (see Degand et al. 2013:1). This issue has been discussed from both formal and functional perspectives. Syntactically modal particles are defined by their position inside the utterance. However, there is little agreement about the functional definition of modal particles.
      Vaskó and Fretheim (1997), for example, define modal particles with regard to their context-adjusting function. In the typical case an illocutionary act modified by a particle 'contains information which the speaker feels that the hearer will not access easily without the speaker's intervention' (p. 253). … 'Regardless of the speech act performed, the speaker's purpose is to make the hearer aware of a particular assumption, or set of assumptions, entertained by the speaker, which the speaker wishes the hearer to accept and to avail himself (sic) of during the conversation' (p. 254). With *nog* the speaker's idea is for example to tell

the hearer to base his/her interpretation on what is likely or probable. Another reason may be to hedge or soften an assertion which may be experienced as brusque.

(3) Men det var <u>nog</u> bara prat. (KE1)
     But that was probably just talk.

Discourse markers such as *actually, well,* and *in fact* are distinguished from modal particles both formally and functionally. They may appear utterance-initially (outside the clause). They typically have functions relating to the sequential organization of discourse for example to mark frames and boundaries in the discourse.

3. Previous research on nog

According to *the Swedish Academy Grammar* (SAG) (Teleman et al. 1999: 117), the Swedish modal particle *nog* (referred to as a modal sentence adverbial in their terminology) has two 'relatively clearly distinct' senses. In the first sense *nog* indicates that the speaker judges the contents of the declarative sentence to be probable (the speaker assumes that…).[1]

(4) Dom vill <u>nog</u> hellre ha Tant Grön, Tant Brun eller något spännande.
     They prefer 'nog' Aunt Green, Aunt Brown or something exciting.

Unlike *väl* (which is hearer-oriented and similar to a tag question) the particle *nog* (in medial position) does not appeal to the hearer for confirmation and is not used in interrogative sentences.

In the second sense *nog* is said to strengthen a speech act. *Nog* guarantees the truth of statements about things which the speaker him/herself is in the position to judge (e.g. reports about perceptions or mental states) or expresses the speaker's commitment to carrying out the action in promises or threats.

---

[1] The examples in this section are from Teleman et al. (1999: 117-118).

(5a)   Jag tycker <u>nog</u> i alla fall som Ulf att det är straffbart.
        I think 'nog' anyhow like Ulf that it is punishable.

The following utterances illustrate *nog* with the function of a promise (5b) or a threat (5c):

(5b)   Jag kan <u>nog</u> både sjunga och spela må ni tro.
        I can 'nog' both play and sing you know.

(5c)   Jag ska <u>nog</u> ge dig, din skurk.
        I will 'nog' kick your butt, you scoundrel.

Nog is reinforcing especially in initial position. In this position the first sense of *nog* (as a probability marker) is not possible:

(6)   <u>Nog</u> fövånar det mig litet att jag fick för pappa och mamma.
        'Nog' surprises it me a little that I was allowed to by my father and mother.

If the utterance with *nog* is used about something the hearer knows better than the speaker, *nog* can have the function of a question or hearer appeal:

(7)   <u>Nog</u> kan jag väl få låna din cykel?
        'Nog' can I borrow 'väl' the bicycle from you?

The reinforcing *nog* is sometimes similar to an adversative adverbial with concessive meaning ('admittedly'):

(8)   <u>Nog</u> vill jag komma ur det här, men jag kan inte.
        'Nog' want I to get out of this, but I cannot.

Teleman et al (1999) propose a number of meanings and functions that *nog* can have in Swedish. By using a parallel corpus where Swedish is one of the compared languages we can check if the functions which have been distinguished have their own translation. The translations can also show meanings or contextual effects of a modal particle which are difficult to discover on the basis of a single language.

4. Material and method

Modal particles are restricted to certain languages. However the function of modal particles can also be translated into languages which do not have modal particles such as English (Waltereit 2001). *Nog* does not have a fixed meaning making it possible to translate it 'uniformly' but has a large repertoire of different meanings depending on the context. The use of translations as evidence for a certain meaning and as a source for contrastive analysis is nothing new. Linguists have often referred to *nog* as 'tests' for different meanings. The method used here is to study the meanings of the modal particle systematically in a corpus of translations.

Parallel corpora for cross-linguistic research of linguistic elements with their translations into the other language have now been available for several decades (see Johansson 2007 for some research based on parallel corpora). They have been a particularly valuable resource to study phenomena in spoken language which do not have a uniform meaning in all contexts (see e.g. the overview in Aijmer and Simon-Vandenbergen 2011 of contrastive corpus studies of pragmatic markers). Studies of Swedish modal particles include Aijmer (1996) and Aijmer (2015) on Swedish *väl*.

The English-Swedish Parallel Corpus (ESPC) contains original texts in English and Swedish with their translations into the other language, altogether 2.8 million words representing both fiction and non-fiction (see Altenberg and Aijmer 2000).[2] The parallel corpus can be used in several ways to establish similarities and differences between languages and to support or modify results based on research on monolingual corpora. Since the ESPC is a bidirectional corpus we can study both the English translations of *nog* in Swedish originals and the English 'sources' of *nog* in the translated texts. I have used the fiction part of the corpus only (about 1.5 million words) since this is the closest correspondence to speech.

---

[2] A description of the corpus is also available at http://www.sprak.gu.se/ english/research/research-activities/corpus-linguistics/corpora-at-the-dll/espc/.

5. The correspondences of nog *in the ESPC*

In this section I will look at the correspondences of *nog* in the fiction part of the ESPC. Altogether there were 142 examples in the translations from Swedish into English. *Nog* was more frequent in original texts than in sources reflecting the fact that modal particles have important discourse functions in the interaction. See Table 1 where the correspondences have been ordered by (total) frequency:

Table 1: Translations and sources of *nog* in the English-Swedish Parallel Corpus

Correspondence	SO→ET	ST←EO	Total
Ø	75 (52.8%)	18 (23.1%)	93 (42.3%)
probably	24 (16.9%)	3 (3.8%)	27 (12.3%)
I suppose	5	3	8
will	3	4	7
of course	6	-	6
I guess	3	3	6
surely	1	5	6
I think	1	5	6
no doubt	5	-	5
must	5	-	5
certainly	4	1	5
I am sure	3	-	3
tag question	1	2	3
might	-	3	3
really	3	-	3
seem to	2	-	2
might well	1	1	2
sure (adverb)	-	2	2
I'm afraid	-	2	2
emphatic *do,*	-	1	1
emphatic *is*	-	1	1
I must say	-	1	1
I dare say	-	1	1
I should think	-	1	1
I reckon	-	1	1
I expect	-	1	1
I suppose…surely	-	1	1
that he is	-	1	1
obviously	-	1	1
would	-	1	1
may	-	1	1
sure enough	-	1	1
negative interrogative question	-	1	1
you'd better	-	1	1
I'd better	-	1	1
just gotta	-	1	1
anyway	-	1	1
in fact	-	1	1
exactly	-	1	1
by the look of them	-	1	1
my advice is	-	1	1
other	1	4	5
Total	**142**	**78**	**220**

Nog has been rendered by many different lexical expressions or constructions. Altogether there were 40 different variants (including zero) only 9 of which were shared by translations and sources. *Of course, no doubt* and *must* were frequent in English translations but not in sources. *Surely* and *I think* were the most frequent variants in sources but they occurred only once in translations.

Omission was the most frequent strategy. In 42.3% of the examples *nog* has either been deleted in English translations or added in the Swedish ones. The translator looks for a correspondence that seems to fit the context or omits *nog* (a zero translation) if the particle is not thought to be important for the message conveyed. *Probably* was the most frequent correspondence (after omission). It was more frequent in English translations than in sources. This suggests that the translator has overused *probably* because of its formal and semantic similarity with *nog*.

The translation paradigm provides us with a messy picture of what *nog* means in different contexts. *Nog* does not seem to have a single meaning but a variety of different functions reflecting its frequency and importance in spoken language. Moreover the English correspondences fall into different word classes (see Table 2). It has been suggested that modal particles (in German) correspond to non-integrated 'formulas' (discourse markers) in English (Fillmore 1984). However in my data the only discourse markers were *anyway* and *in fact* (both occurring only once). The most frequent type of correspondence was instead a modal adverb. The functions of *nog* are also similar to those associated with 'modal tags' such as *I think, I suppose* expressing a degree of certainty (cf. de Haan 2006: 38 for the term 'modal tag'). *Nog* can also correspond to a modal auxiliary (in particular *will* and *must*). On the other hand *nog* was rarely translated with a tag question.

Table 2: Word class correspondences of *nog* in English

Modal adverb	probably, of course, surely, no doubt, certainly, really, sure, obviously, just (gotta)	53
Modal tag	I think, I guess, I should think, I suppose, I'm sure, I'm afraid, I must say, I dare say, I expect	30
Modal auxiliary	must, will, might, might well, may, would, could; would rather, had better	17
Tag questions	e.g. isn't it, wouldn't I	3
Discourse marker	in fact, anyway	2
Modal verb	seem	2
Other modal expressions	e.g. my advice is, by the look of them, that he is, negative interrogative sentence	5

6. Functions of nog

Modal particles are multifunctional and context-bound. They therefore raise a number of questions about semantics and pragmatics. Do they have one meaning or should we account for their multifunctionality in a polysemous approach? *Nog* signals in principle that the speaker has sufficient grounds for the truth of the utterance whether this involves that something is probable or true (Borgstam 1977; Lindström 2008:98). It follows that *nog* is subjective or speaker-oriented (Teleman et al 1999; Fretheim 1981). There is no appeal to the hearer or someone who knows better but the speaker takes full responsibility for his/her attitudes and actions.

  Nog seems to be associated with different core aspects depending on its position and the presence of stress. The following discussion is therefore organized in two parts: the first examines unstressed *nog* placed medially and the second example where stressed *nog* is placed initially.

6.1 Nog *in medial position*

6.1.1. Nog *indicating probability*
The translation with *probably* signals that *nog* expresses a high degree of certainty (probability):

(10) — Men det är <u>nog</u> bara ljummet, sa Pretorius. (ARP1)[3]
     "Yes, but it 's <u>probably</u> got rather cold," said Pretorius. (ARP1T)

The speaker makes a fairly confident statement about the temperature of the coffee.
    In (11) *nog* interacts with *I suspect* in the main clause. The translator has used *probably:*

(11) - Då misstänker jag att hon <u>nog</u> kommer att bli överraskad. (HM1)
     "Then I suspect she 'll <u>probably</u> be surprised." (HM1T)

    The probability meaning seems to be the most frequent one of *nog* in translations into British English. The meaning can also be expressed by other means than *probably.*
    No doubt expresses probability rather than certainty and has the same meaning as *probably.* According to Quirk et al. (1985: 623) 'it in fact implies some doubt and is synonymous with "very probably"':

(12) De skulle <u>nog</u> klara sitt husköp. (HM2)
     They would <u>no doubt</u> cope with buying their house. (HM2T)

    Probability can also be expressed by a modal auxiliary (13). *Will* as a translation indicates that the speaker has sufficient knowledge of the facts to judge that something is true:

(13) Jojo, tänkte fastern. Julgransplundringen är <u>nog</u> snart över om de fortsätter på det här viset. (ARP1)
     "Oh, yes," thought Auntie, "presents under the Christmas tree <u>will</u> soon be a thing of the past if this goes on. (ARP1T)

    In (14) the translation with *must* suggests that the speaker has enough background information for inferring that 'he is about fifty':

(14) Han är <u>nog</u> omkring femtio — fast han ser yngre ut, tänker jag och ser in i hans blick. (MS1)

---

[3] For information about the text codes, see Altenberg and Aijmer (2000).

He <u>must be</u> around fifty but looks younger, I think, and look into his eyes. (MS1T)

In (15) *nog* refers to the facts or circumstances on the basis of which the speaker makes a judgement. The translator has used 'by the look of them':

(15)   Lögnen kom av sig själv och de accepterade den utan vidare. Ingen av dem sa någonting efter det. Valerie satt och stirrade på deras kortklippta grå nackar och förarens åldersfläckiga händer på ratten. De är <u>nog</u> systrar, tänkte hon. (PDJ1T)
The lie came easily to her and was as easily accepted. Nothing more was said by any of them. She sat looking at the backs of the two grey, cropped heads, watching the driver's age-speckled hands on the wheel. Sisters, she thought, <u>by the look of them</u>. (PDJ1)

6.1.2. Nog *as a downtoner*
The speaker may also use *nog* for reasons having to do with politeness. When *nog* has been translated by a modal tag such as *I suppose, I guess, I dare say, I should think* its function is to soften or tone down the illocutionary force or more generally the speaker's commitment to what is said:

(16)   Sen är jag <u>nog</u> inte lätt att umgås med eftersom jag till stora delar består av sällskapliga gäster i det där hotellet jag talade om. (RJ1)
And <u>I suppose</u> I 'm not that easy to get on with because I 've got this hotel full of guests that I mentioned earlier. (RJ1T)

In (17) the speaker has used *I'm afraid* apologizing for not being able to answer the question. The translator's *nog* makes the answer less abrupt than the corresponding utterance without a modal particle.

(17)   "I 'm afraid I ca n't answer that, Mr Orloff," said Cooper honestly. (MW1)
"Jag kan <u>nog</u> inte svara på den frågan, herr Orloff", sade Cooper ärligt. (MW1T)

In the examples quoted *nog* is associated with softening an illocutionary act. The same sentence without *nog* would sound impositive and brusque and threaten the hearer's negative face (the hearer's want to have his/her freedom of action unimpeded) (Brown and Levinson 1978).

6.1.2 Nog *indicating certainty*

In (18) *nog* seems to suggest 'complete certainty' as indicated by the translator's *certainly* (Aijmer 2002):

(18)   Det var <u>nog</u> kallare ute än det verkade. Röken bolmade ur skorstenarna och slog ner. Det blåste och röken svepte i vita slöjor förbi hennes fönster. (MG1)
It <u>certainly</u> was colder outside than it seemed. It was windy, and the smoke from the chimneys swept by her window in white veils. The sun broke through the clouds, gilding the veils of smoke. (MG1T)

However, when *nog* is translated as 'certainly' or another expression of certainty the reason may also be that the speaker needs some reassurance that the grounds for judging something to be true are sufficient (cf. Solberg 1990: 55 who considers one of the most important elements of the Norwegian particle *nok* (cognate with Swedish *nog*) to be that the speaker tries to prove to herself that something is the case). In (19) the translator has used 'I'm sure' to express the speaker's confident prediction or self-assurance that something will be the case. This interpretation is further strengthened by the addition of an explanation for the speaker's (un)certainty (H B may be glad to spend some time chatting since she is alone).

(19)   Fast om söndagarna tog hon sig ändå för det mesta åt Storholmträsk, Gammlundström brukade skjutsa henne, han sade: jag får <u>nog</u> en kaffetår av Hanna Burvall, hon kan vara glad att få prata bort en stund, hon är ju som ensammen. (TL1)
Though on most Sundays she got herself to Storholmsträsk. Old Lundström used to drive her there, he said: "I 'm sure I 'll get a cup of coffee from Hanna Burvall. She may be glad to spend some time chatting. She is as it were alone. (TL1T)

6.1.3 Emotional and subjective uses
As the Swedish Academy Grammar (Teleman et al. 1999) has pointed out, *nog* has subjective or affective uses in addition to the hedging or strengthening (epistemic) meaning. *Nog* is for example used as a modifier of the verb in subjective statements (statements about the speaker's own perceptions, needs or mental states).

In (20) the equivalent in English is the emphatic *do*:

(20) So you see, I <u>do</u> need a consciousness-raising group after all. (MD1)
Du ser att jag behöver <u>nog</u> gå i gruppterapi i alla fall. (MD1T)

However *nog* has often been omitted by the translator after verbs expressing a subjective opinion or mental state (*tror jag nog* 'think I nog'*, jag märkte nog* 'I noticed nog', *jag vet nog* 'I know nog'). In (21) *I know* has been regarded as sufficient in the English translation:

(21) Jag vet <u>nog</u> att man måste jobba för Ödet. (RJ1)
I know you have to work for Fate. (RJ1T)

In (22) and (23) *nog* has been added in the Swedish translations after *tycka* and *tro* to make the Swedish text more idiomatic:

(22) "I think you 'd best," he said. "What would your name be?" (SG1)
"Ja, det tycker <u>nog</u> jag också", sade han. "Hur var ert namn?" (SG1T)

Tror jag nog can be regarded as an idiomatic expression which is typically used in response utterances.

(23) Do you think you can find it?"
"Got a tongue in my head," he said, peering at the maze of roads.
"Reckon so." (DF1)
Tror du att du kan hitta den?"
"Kan ju fråga mig fram", sa han och kikade på villervallan av vägar.
"<u>Tror jag nog</u>." (DF1T)

Nog is needed to tone down a too blunt subjective statement. English has no explicit correspondence in such cases.

6.1.4 Nog *with deontic uses*
Nog can be interpreted as deontic in combination with a deontic modal auxiliary. The 'deontic *nog*' expresses the speaker's attitude towards a possible action with regard to whether it is appropriate or morally right (cf. Palmer 1986: 120). In (15) the source text has *had better* 'a comparative modal' (van der Auwera et al 2013) used to express advice given by the speaker to the hearer (the hearer should do something or it is 'best' for the hearer to do something). The translator has used *få nog* ('may nog', 'must nog'):

(24)   You 'd better keep my dinner warm." (DL1)
       Du får nog hålla maten varm." (DL1T)

(*Ni*) *ska* ('you shall') combines with *nog* and is (also) used to give advice (the speaker judges an action to be the best one). The source text has 'my advice is':

(25)   "I 'll see what I can find out for you back at the lab, but my advice is, do n't hold your breath. (MW1)
       "Jag ska se vad jag kan hitta åt er när jag kommer till labbet, men ni ska nog inte ha för stora förhoppningar. (MW1T)

   In (26) *nog* is a part of the Swedish modal idiom *det är bäst att* ('it is best that'). The source text has *it's time* (you did something):

(26)   "You 're raving ever so slightly, darling, and it 's time you went home." (RDA1)
       Det är nog bäst att du går hem." (RDA1T)

In (27) the English original has *just gotta*. *Nog* as a translation of *just* emphasizes what ought to be done:

(27) "Sometimes, you <u>just</u> gotta stay there and teach him how to go through the bad and good of whatever comes." (GN1)
Någon gång blir du <u>nog</u> tvungen att stanna kvar och lära honom att ta ont och gott som det kommer. " (GN1T)

6.1.5 Nog *with the meaning of a promise or a threat*
Nog can signal that the utterance should be taken as a threat (*jag ska nog* 'I shall nog'). The particle reinforces (or mitigates) the speaker's attitude (the speaker commits herself to a future action). In the English translation *nog* has been omitted:

(28) "Ät gröt, om du kan tugga den, och lämna vildoxarna åt mej", sa Mattis.
"Dem <u>ska jag nog</u> bli färdig med när tiden är inne." (AL1)
"Eat your porridge, if you can chew it, and leave wild bulls to me," said Matt. "<u>I 'll deal with them</u> when the time comes." (AL1T)

    I'm going to get you in the English original text is typically a threat. In the Swedish translation *nog* conveys the speaker's emotional stance:

(29) — <u>I 'm going to</u> get you, I told Edward Swanwick. (RDO1)
      — Du ska <u>nog</u> få igen, sa jag till Edward Swanwick. (RDO1T)

Nog is associated with promises in the idiomatic expression *det är/blir… bra*. The translator has used *be going to* to express the speaker's emotional involvement in the activity and its effects on the hearer. The speaker offers the hearer comfort and reassurance that things will be all right.

(30) — Det <u>blir nog</u> bra ska du se. (GT1)
      "It <u>'s going to</u> be all right, you 'll see." (GT1T)

6.2 Nog *in initial position*
Roughly 17% of the examples of *nog* had initial position. In initial position *nog* was always stressed. The translation with *really* makes it clear that it can mean emphasis:

(31) Åjo, <u>nog</u> är det ganska tyst och tomt på landet nu för tiden.
(SC1)
"Oh yes, it <u>really</u> is silent and empty in the country nowadays.
(SC1T)

In (32) *certainly* is used in the source text and the translator has used
nog. Certainly is not neutral but implies an appeal for confirmation:

(32) He had no idea when she had given that dress up, but <u>certainly</u> it
was years and years ago. (AT1)
Han hade ingen aning om när hon hade gjort sig av med den där
klänningen, men <u>nog</u> var det för flera år sedan.(AT1T)

Both *väl* and *nog* can be used to appeal to the hearer for positive
feedback but in different ways. Teleman et al (1999: 116) describe the
meaning of *väl* as follows, 'In conversation a declarative sentence with
väl can be used as a careful question, an appeal to the hearer for
agreement' (translation KA)'. The initial *nog* on the other hand is a
convenient way of introducing 'contradictory assumptions' without
being explicit about which these assumptions are.
   Further evidence for the hearer-oriented function of intial *nog* is
found in (33) where the English source text contains a negative
interrogative sentence. These interrogatives are special in being biased
towards a particular interpretation of the answer to the question.
According to Biber et al (1999: 1114), 'negative interrogatives challenge
a negative expectation that has been assumed to exist in the context, and
thus indicate the speaker's inclination towards a positive response':

(33) She would open the front door and smell home.
She would pass through the rooms where she 'd been so happy all
these years. (<u>Had n't she</u> been happy?) She would find the cat
stretched out on the couch, long and lazy and languid, and she 'd
settle on the cushion next to her and think, How could I have left?
(AT1)
Hon skulle öppna dörren och känna att det luktade hemma.
Hon skulle gå genom de rum där hon hade varit så lycklig alla
dessa år. (<u>Nog hade hon väl</u> varit lycklig?) Hon skulle få se katten
ligga utsträckt på dyschan, lång och lat och dåsig, och hon skulle

sätta sig bredvid den och tänka: Hur kunde jag ge mig iväg härifrån? (AT1T)

Both *nog* and *väl* have the function to appeal to the hearer for (positive) feedback. *Nog* conveys that the speaker tries to convince herself that she had been happy at the same time as she is invoking the scenario of not being happy. *Nog* (but not *väl*) can also convey emotions such as resignation or reluctant admission ('she should have been happy shouldn't she') associated with what ought to be the case.

  Nog can also correspond to a tag question in English. The sentence with a tag question presupposes a positive response. In (34) the speaker appeals to the hearer to share her assumption that it would be interesting to study the parents (to find out why they forced their daughters to marry to preserve their respectability).

(34) It 's interesting, <u>is n't it</u>, to observe the parents. It would be quite wrong to say that they sold their daughter to preserve their respectability; they would n't have done that. (RDA1)
Men <u>nog</u> är det intressant att studera föräldrarna. Det skulle vara helt fel att säga att de sålde sin dotter för att bevara hedern, det hade de aldrig gjort. (RDA1T)

In (35) the English original contains *surely*. The translation with (*nog*) *borde* ('should', 'ought to') is evidence that the question has a deontic bias: the senator should have protested a little more:

(35) <u>SURELY</u> THE SENATOR might have argued a little more, said the Daimon Maimas. (RDA1)
— <u>NOG</u> BORDE VÄL senatorn ha protesterat lite mer, sa daimon Maimas (RDA1T)

  Another context-bound meaning of *nog* is concessive as indicated by translations such as *of course, obviously*. In (36) the clause containing *nog* is followed by a *but*-clause which foregrounds a more convincing argument so that the first clause gets a concessive function ('admittedly'):

(36) <u>*Nog*</u> *kunde hon springa som en räv och nog visste hon skogens alla gömmen,* <u>*men*</u> *vittrorna kom envist efter henne, och hon hörde deras gälla skrik:* (AL1)
<u>Of course</u> she could run like a fox, and of course she knew every hiding place in the forest, <u>but</u> the harpies pursued her stubbornly, and she heard their strident cries, "Ho, ho, pretty little human, blood will run now, ho, ho!" (AL1T)

To sum up, *nog* in initial position has translations signalling meanings such as emphasis (*really*), contradictory assumptions and hearer appeal (*certainly, surely*, negative interrogatives, tag question) or concessive meaning (*of course*). In the hearer-appealing function it could have deontic implications about what ought to be the case. Some possible contextual effects of using *nog* are resignation and reluctance to accept that something is the case.

7. Conclusion

The key to the use of *nog* seems to be that the speaker assumes sole responsibility for the truth of his/her utterances, attitudes and opinions, and actions. This is compatible with the general meaning that the speaker has sufficient for judging that something is true. However, *nog* does not have a single meaning but seems to be what Norén and Linell (2006: 12) refer to as 'relatively polysemous'.

Probability was one of the most frequent meanings of *nog* in medial position. *Nog* can also mean (almost) certainty. When *nog* corresponded to *certainly* or *I am sure* the speaker looks for reassurance that the grounds for judging something to be true are sufficient.

An important factor accounting for the uses of *nog* in medial position is politeness. *Nog* does not express certainty but functions as a 'downtoner' softening an opinion or an utterance which might be offensive to the hearer unless mitigated. *Nog* was also used in combinations with mental verbs such as *jag vet nog* (I know 'nog') or *jag tycker nog* (I think 'nog') especially in response utterances with a softening effect. *Nog* had no correspondence in English in such cases.

The translations with a deontic modal auxiliary (e.g. *had better*) have indicated another semantic element in the analysis of *nog*; i.e. 'what ought to be the case' or 'what you have to do' according to some moral principle. The deontic meaning was particularly clear when *nog* was used

with modal auxiliaries to give advice, recommendations or suggestions (e.g. *du ska nog* 'you shall nog'; compare also *det är nog bäst* 'it is nog best').

Nog was also found in 'commissive' speech acts such as threats and promises. The speaker takes upon him/herself to carry out an action which is either favourable or unfavourable to the hearer.

Nog, when initial, often involved contradictory assumptions. Contrasting opinions account for example for the translations of a sentence with initial *nog* by a negative interrogative sentence (or a declarative sentence with a tag question). The meaning of *nog* comes close to *väl* ('I suppose') and hearer-appeal. The initial *nog* can also be emphatic or used concessively to prepare the hearer for an objection in a following *but*-clause.

Modal particles need to be studied both in monolingual corpora and contrastively. The study of *nog* on the basis of the English-Swedish Parallel corpus has shown that the contrastive perspective can enrich analyses based on a single language only. However we need to study many more modal particles (and groups of modal particles) in different language pairs to get a better picture of how languages structure a particular pragmatic-semantic field both by means of modal particles and in other ways.

References

Aijmer, K. 1996. Swedish modal particles in a contrastive perspective. *Language Sciences* 18(1-2): 393-427.

Aijmer, K. 2002. Modal adverbs of certainty and uncertainty in an English-Swedish perspective. In Hasselgård, H., Johansson, S., Behrens, B. and C. Fabricius-Hansen (eds), *Information structure in a cross-linguistic perspective.* Amsterdam/New York: Rodopi. 97-112.

Aijmer, K. 2015. The Swedish modal particle *väl* in a contrastive perspective. In Aijmer, K. and H. Hasselgård (eds), *Cross-linguistic studies at the interface between lexis and grammar.* Special issue of *Nordic Journal of English Studies* 14(1): 174-200.

Aijmer, K. and A.-M. Simon-Vandenbergen 2011. Pragmatic markers. In Zienkowski, J., Östman, J.-O. and J. Verschueren (eds), *Discursive Pragmatics.* Amsterdam/Philadelphia: John Benjamins. 223-247

Altenberg, B. and K. Aijmer. 2001. The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies. In Mair, C. and M. Hundt (eds), *Corpus linguistics and linguistic theory. Papers from the 20th International Conference on English Language Research on Computerized Corpora* (*ICAME 20*) *Freiburg im Breisgau 1999*, 15-33. Amsterdam & Philadelphia: Rodopi.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *The Longman grammar of spoken and written English.* London: Longman.

Borgstam, S. 1977. Nog är tillräckligt. *Nysvenska studier* 57: 217-226.

Borthen, K. and Knudsen, C. S. 2014. Semantiske og pragmatiske aspeker vetd trykklett *visst. Norsk Lingvistisk Tidsskrift* 31:360-411.

Brown, P. and S. Levinson. 1978. Universals in language usage. Politeness phenomena. In Goody, E. (ed.), *Questions and politeness: Strategies in social interaction.* Cambridge: Cambridge University Press. 56-310.

Cuenca, M. J. 2013. The fuzzy boundaries between discourse marking and modal marking. In Degand, L., B. Cornillie and P. Pietandrea (eds), *Discourse markers and modal particles. Categorization and description.* Amsterdam and Philadelphia: John Benjamins. 191-216.

Davidsen-Nielsen, N. 1996. Discourse particles in Danish, In Engberg-Pedersen, E., Fortescue, M., Harder, P., Heltoft, L. and L. Falster-Jacobsen (eds), *Content, expression and structure. Studies in Danish functional grammar.* Amsterdam/Philadelphia: John Benjamins. 283-314.

Degand, L., B. Cornillie and P. Pietandrea (eds). 2013. Modal particles and discourse markers: Two sides of the same coin. In Degand, L., Cornillie, B., and P. Pietandrea (eds), *Discourse markers and modal particles. Categorization and description.* Amsterdam and Philadelphia: John Benjamins. 1-18.

Diewald, G. 2006. Discourse particles and modal particles as grammatical elements. In Fischer, K. (ed.), *Approaches to discourse particles.* Amsterdam: Elsevier. 403-425.

Diewald, G. 2013. '*Same same but different'* —Modal particles, discourse markers and the art (and purpose) of categorization. In Degand, L., Cornillie, B., and P. Pietandrea (eds), *Discourse markers*

and modal particles. Categorization and description. Amsterdam and Philadelphia: John Benjamins. 19-45.

Fillmore, C. 1984. Remarks on contrastive pragmatics. In Fisiak, J. *Contrastive linguistics.* Berlin: Mouton de Gruyter. 120-141.

Fretheim, T. 1981.'Ego'-dempere og 'alter'-dempere. *Maal og Minne,* 86-100.

de Haan, F. 2005.Typological approaches to modality. In Frawley. W. (ed.) with the assistance of Eschenroeder, R.; Mills, S.; Nguyen, T. *The expression of modality.* Berlin/New York: Mouton de Gruyter. 27-69.

Johansson, S. 2007. *Seeing through multilingual corpora. On the use of corpora in contrastive studies.* Amsterdam/Philadelphia: John Benjamins.

Lindström, J. 2008. *Tur och ordning. Introduktion till svensk samtalsgrammatik.* Stockholm: Norstedts akademiska förlag.

Norén, K. and P. Linell. 2006. Meningspotentialer i den språkliga praktiken. Meddelanden från Institutionen för svenska språket (MISS). Göteborgs universitet.

Palmer, F.R. 1986. *Mood and modality.* Cambridge: Cambridge University Press.

Quirk, R, S. Greenbaum, G. Leech and J. Svartvik. 1985. *A comprehensive grammar of English.* London: Longman.

Solberg, T.K. 1990. *Modalpartikler i norsk.* Hovedoppgave ved Avdeling for nordisk språk og litteratur. Institutt for nordistikk og litteraturvitenskap. Universitetet i Oslo.

Teleman, U., S. Hellberg and E. Andersson. 1999. *Svenska Akademiens grammatik.* Band 4. Stockholm: Norstedt.

van der Auwera, J., D. Noël and A. Van linden. 2013. *Had better, 'd better* and *better*: diachronic and transatlantic variation". In Marín-Arrese, J.I., M. Carretero, J. Arús Hita and J. van der Auwera (eds), *English modality: core, periphery and evidentiality.* Berlin: Mouton de Gruyter, 119-153.

Vaskó, I. and T. Fretheim. 2007. Some central pragmatic functions of the Norwegian particles *altså* and *nemlig.* In Swan, T. and O.J. Westvik (eds), *Modality in Germanic languages. Historical and comparative perspectives.* Berlin: Mouton de Gruyter. 233-292.

Waltereit, R. 2001. Modal particles and their functional equivalents: A speech-act-theoretic approach. *Journal of Pragmatics* 33: 1391-1417.

Swedish *hinna* viewed through its English correspondences – *have time* or *be able to?*

Mats Johansson and Lene Nordrum, Lund University

Abstract
This article considers the Swedish verb *hinna* and its English correspondences in the English-Swedish Parallel Corpus – both in the direction Swedish original to English translation and in the direction English original to Swedish translation. Swedish *hinna* lacks a straightforward equivalent in English; thus its English correspondences in parallel corpora can shed light on its meaning and use, as described in the method 'seeing language through multilingual corpora' (Johansson 2007). The results for *hinna* suggest that *hinna* is monosemous, but that its core meaning is enriched by implied ones (presupposition and conversational implicature). Based on the observed translations and sources of *hinna*, we propose that in the majority of our data *hinna* denotes time sufficiency and also presupposes ability and conversationally implicates actualization. We relate our discussion of *hinna* to van der Auwera and Plungian's (1998) semantic map of modality, in which ability and sufficiency are seen as contiguous modal meanings.[1]

1 Introduction

The Swedish semi-auxiliary verb *hinna*, as in (1), lacks a straightforward equivalent in English, which makes it a good candidate for investigation in parallel corpora, both from a purely contrastive point of view and for the semantic information about *hinna* offered by the English correspondences.

(1)   Jag har inte hunnit   äta (ännu).
       I   have not HINNA.ptc eat (yet).
       'I haven't had the/enough time to eat (yet)'

In the major Swedish reference grammar, *Svenska Akademiens Grammatik* (SAG) (Teleman et al 1999), *hinna* as in (1) is defined as a semi-auxiliary in the category *potential modal auxiliary*. Teleman et al (1999(4): 297) describe the meaning of *hinna* as marking the ability of the subject referent to realize a particular action within a relevant time frame, and also state that *hinna* typically requires an animate, agentive Subject. In addition, they argue that there is usually no clear meaning difference between modal auxiliary verbs and lexical verbs with the same form in Swedish (Teleman et al 1999(4): 282), but do not comment specifically on *hinna* in this respect. Aijmer (2004), in a study of "verbs and related constructions with the meaning dynamic possibility or ability in English and Swedish and their relation to actuality", describes *hinna* as a verb making the "enabling conditions" of someone's ability explicit. In our terms, the enabling condition for *hinna* corresponds to time sufficiency, i.e. the availability of enough time for some event to take place.

What is interesting about *hinna*, then, is that it seems to communicate **both** a notion of *time sufficiency* and a notion of *ability*, both of which we take to be modal meanings. Thus, *hinna* can be glossed roughly as 'have enough time to be able to [do sth]'. Such observations about the semantic properties of *hinna* have not been studied extensively in light of Swedish corpus data and there is no corpus-based study of *hinna*'s correspondences in English.

The aim of this paper is therefore to seek to account for semantic properties of the Swedish verb *hinna* as mirrored in its English correspondences in the English-Swedish Parallel Corpus (ESPC) (cf. Johansson 2007). The data consist of *hinna* and its English correspondences retrieved from Swedish original texts (fictional and non-fictional) translated into English (SO→ET) and English original texts translated into Swedish (ST←EO). One main advantage of authentic texts and their translations is that neither the source texts nor the translations were produced with a theoretical agenda (cf. Aijmer 2004), which should provide a richer spectrum of contexts for *hinna* compared to constructed examples. For a thorough description of the ESPC, we refer to Altenberg and Aijmer (2000).[2]

---

[2] A description of the corpus is also available at
http://www.sprak.gu.se/english/research/research-activities/corpus-linguistics/corpora-at-the-dll/espc/.

A point of departure for the study is that if, as we claim, *hinna* carries multiple simultaneous meanings, they may all be made more visible by the English correspondences – both the sources and the translations. As pointed out by Johansson (2007: 28) "one of the most fascinating aspects of multilingual corpora is that they can make meanings visible through translation. Ambiguity and vagueness are revealed through translation patterns." In addition, we aim to consider whether implications (presuppositions and implicatures) in original texts can show up in the correspondences not only as presuppositions and implicatures, but also as as explicit assertions or entailments[3]. We thus follow Chesterman (1998) in regarding *the tertium comparationis* of translations not as a basis for the analysis but as its result. In traditional contrastive analysis terms, we approach the parallel corpus data from the point of view of the single verb *hinna*, rather than from the point of view of a tertium comparationis, i.e. some independent meaning component which could form the basis of, say, a comparison of Swedish and English expressions of time sufficiency and ability.

We argue that *hinna* in its most common semi-auxiliary uses involves three meaning components: time sufficiency, ability and actuality*,* of which only time sufficiency is part of the core semantics of *hinna*, ability is presupposed and actuality is conversationally implicated. We show that the congruent English correspondences typically reflect one or two of these components, time sufficiency (e.g. *have time*), ability (e.g. *can*, *be able to*) or actuality (*manage*), but that the complex meaning of *hinna* is mirrored in how these congruent correspondences interact with their context. For example, what is (lexically) presupposed by *hinna* may be asserted or presupposed (or otherwise implied) by an English correspondence (such as e.g. *can*) in conjunction with the surrounding context.

The article is organized as follows. In section 2, we take a first look at the English correspondences of *hinna*, and provide a rough classification of *congruent* correspondences in terms of the meaning components they reflect. Section 3 takes a closer look at the meaning correspondences outlined in section 2, time sufficiency, ability and actuality. Section 4 shows the co-occurrence of time sufficiency and

---

[3] Our primary aim is not to propose a novel approach to these notions. Instead, we employ what we take to be reasonably standard characterizations of entailment, presupposition and implicature, and define them where it is required.

ability/actuality in the translation examples. In section 5, we re-examine our data and take a closer look at zero translation and show that in a majority of cases, time sufficiency is contextually present, e.g. in temporal clauses. Section 6 further motivates our suggestion that time sufficiency constitutes the core meaning of *hinna*, and section 7 embeds our discussion in a semantic map of modality, where ability and sufficiency are seen as contiguous modal meanings. Section 8 concludes.

2. Hinna *and its correspondences in the directions SO→ET and ST←EO*
In this section, we take a 'surface' look at the English correspondences of *hinna* in the ESPC to explore whether this material can tell us something about how the meaning components of *hinna* can be described. The surface approach means that we only consider the *congruent* English translation correspondences of *hinna* – i.e. those correspondences where *hinna* is translated with a verb construction and the source syntax is kept intact (Johansson 2007: 23-26). Table 1 gives the English translations and sources of *hinna*. The type of meaning realized by the correspondence is specified in the column 'semantic type'.

Table 1 English sources and translations of *hinna* in the ESPC

Correspondence	SO→ET		ST←EO		Semantic type
	N	%	N	%	
Ø	27	28	34	39.5	?
have time	17	17.6	7	8.1	Time sufficiency
manage	15	15.6	-	0	Actuality
can	11	10.6	13	15.1	Ability
get there/as far as	8	7.7	7	8.1	?
catch up to/with	6	5.7	6	6.9	?
be able to	4		1		Ability
be time	2		1		Time sufficiency
find (the) time	2		-		Time sufficiency
make it	2		-		Actuality
arrive	1		3		?
reach	1		-		?
be capable of	1		-		Ability
chase after	-		1		?
complete	1		-		?
forestall	-		1		?
gain on	-		1		?
get around to	-		1		?
have a chance to	-		1		?
other	6		8		?
Total	104	100	86	100	

* Column one gives the verb phrases that show up as English translations of *hinna* in Swedish original texts (SO→ET) or as the source of *hinna* in Swedish translations from English original texts (ST←EO)*;* columns two and three give the raw frequencies of these – with percentage information for the most common correspondences, and column four our classification of the translation correspondence into time sufficiency, ability or actuality meaning. The translation correspondences that are not easily categorized are marked with a question mark in column four.

A first glance at Table 1 suggests that *hinna* is polysemous since the majority of the congruent correspondences seem to reflect *one* of the time sufficiency, ability and actuality meanings. In addition, Table 1 shows a third meaning, indicated by a question mark. The correspondences of this type are, in order of frequency, *get there/as far as, catch up to/with, arrive, reach, chase after, complete, forestall, gain on*. Common for these correspondences is that they mirror a construction with [*hinna + place adverb*] and thus *only* correspond to *hinna* as a main verb, as illustrated in (2):

(2)   Redan innan han hade <u>hunnit fram</u> började han tala lugnande till
        sig själv, inåtvänt och utan att röra på läpparna. (KOB1T)
        Even before he <u>got there,</u> he had started talking to himself
        Inwardly, without moving his lips, in order to cool down.
        (KOB1T)

We tentatively suggest that correspondences such as (2) mirror an
accomplishment meaning involving 'getting somewhere' or 'reaching an
end-point', and that a possible explanation for why this meaning only
shows up for main verbs is that it reflects a historically older and less
modal meaning than the other correspondences. According to *Svenska
Akademiens Ordbok* (SAOB) (http://g3.spraakdata.gu.se/saob/), *hinna* is
historically related to Gothic *hinþan* ('catch') and Old High German
heriunda ('prey') as well as English *hunt,* and in Old Norse it partly
merged with *inna* ('reach an end', 'finish'). The meaning in (2) seems
close to this historical origin and can be glossed roughly as ''someone
succeeds in getting to a certain end point''. Importantly, though, ability is
always entailed in these cases – if you reach somewhere or get
somewhere, it follows that you have the ability to do so. Moreover, time
sufficiency may also be involved. In examples such as (2), for instance,
time sufficiency meaning is present in the sense that the example is a
comment on the brevity of the time period it took for someone to get
somehwere (i.e. not much time had passed before he started talking to
himself). Nevertheless, since the 'accomplishment' correspondence is the
only type that is not reflected in both lexical- and auxiliary use of *hinna*,
we will largely leave it aside in our discussion.

Unlike the 'accomplishment' correspondences, the correspondences
marked as time sufficiency, ability and actuality in Table 1 support the
observation made by Teleman et al (1999(4): 282) that there is no sharp
demarcation line between auxiliary and lexical (main verb) uses of verbs
such as *hinna*. The absence of a clear distinction is illustrated by cases
like (3-8) where we have the same translations (*have time* in (3-4), *can* in
(5-6) and *manage* in (7-8)) regardless of the status of *hinna* as a lexical
(3, 5 and 7) or auxiliary verb (4, 6 and 8):

(3)   Men de <u>hann</u> inte ut. (KE1)
        But they <u>had n't time</u>. (KE1T)

(4)   Innan den <u>hunnit</u> samla sig trampade farmor rätt in i den: hon gav sig inte. (AP1T)
Before it <u>had time</u> to collect itself, Gran was pedalling right into it. (AP1T)

(5)   Hon kväljdes åt grymheten och sprang för galenskapen eller hukade sig om hon inte <u>hann</u> undan. (KE2)
Cruelty turned her stomach and she fled from folly, or ducked if she <u>couldn't</u> escape. (KE2T)

(6)   Hans fingertoppar rör vid mitt ansikte innan jag <u>hinner</u> dra mig undan. (MS1)
His fingertips touch my face before I <u>can</u> retreat. (MS1T)

(7)   Sen kommer gammelfrun, pigg och rosig och säger att hon har gjort smörgås, det är tio tusen räkmackor och kaffe och sen går Olle och jag runt runt på gårdsplanen tills det blir middag. Vi <u>hinner</u> fyra varv. (IU1)
His wife, lively and rosy cheeked, said she had made a few smörgås — seemingly about ten thousand of them, with prawns — and then Olle and I tried to walk round the place until dinner time. We <u>managed</u> four circuits. (IU1)

(8)   "Ja, jämmer och död, den ligger nu på andra sidan avgrunden, och det blir synd om den som inte kan hålla sej, tills vi <u>hinner</u> inrätta en ny." (AL1)
Yes, death and destruction! It's on the other side of the chasm now, and I 'm sorry for anyone who ca n't contain himself until we <u>manage</u> to build a new one." (AL1T)

In the remainder of the paper, we focus on the time sufficiency/ability/actuality correspondences, predominantly in auxiliary- or semi-auxiliary functions and only briefly return to the accomplishment meaning towards the end of the paper.

3. Time sufficiency, ability, and actualization
In this section, we discuss the English correspondences of *hinna* intitially classified as time sufficiency, ability and actuality in more detail. Section 3.1 deals with time sufficiency, 3.2 with ability and 3.3 with actuality.

3.1 Time sufficiency (as participant-external possibility)
In our material (see Table 1), the time sufficiency meaning of *hinna* is most clearly reflected in the correspondences with *have time*, where time sufficiency is conceptualized as time possession. Example (9) and (10) are illustrations:

(9)   Hon hade inte <u>hunnit</u> bli riktigt vuxen när hon fick mig. (AP1)
        She had n't really <u>had time</u> to become grown up when she had had me. (AP1T)

(10)   She <u>had just enough time</u> to rush in, pick him up, see if he was wet or marked in some way, and then go back to work. (GN1)
        Hon <u>hann bara precis</u> springa in, lyfta upp honom och titta efter om han var våt eller skadad på något sätt och sedan skynda sig tillbaka till arbetet. (GN1T)

In our view, the time sufficiency translations reflect the core semantics of *hinna*. The assumption is that the time-sufficiency notion is always present in *hinna* and therefore part of a Swedish native speaker's knowledge of the word *hinna*. One indication of this core status is that *ha tillräckligt med tid* 'have enough time' seems to be the most natural paraphrase of *hinna*, as also indicated by Teleman et al's (1999) gloss, referred to in section 1, above.

Thus, it follows that *ha tillräckligt med tid* can be understood as an entailment of *hinna*. Generally, the criterion for entailment is that in a sentence pair such as (11ab), (11b) is an entailment of (11a) if the sentences are related in such a way that if (11a) is true, (11b) is also true – irrespective of context. Further, typical of entailments is that they do not hold under negation, an observation which holds true for sentence pair (11ab): If (11a) is negated, (11b) is no longer an entailment. It is possible that (11b) is true, but we cannot know without context. If it is true, the interpretation would probably be that Paul had sufficient time to do it, but failed to because his abilites were subnormal.

(11)   a. Paul hann göra det.
          Paul HINNA.past do it.
          'Paul had (sufficient) time to do it.'

        b. Paul hade tillräckligt med tid för att göra det.
          'Paul had sufficient time to do it.'

Some additional observations about sufficiency are worth noting. For one, sufficiency is most naturally understood as degree in relation to a scale or a position on a scale. The kind of sufficiency relevant for *hinna* is a relation between time periods. For instance, to guarantee that you have sufficient time to drive to work, you estimate how long the drive will take and relate that time to the period between the time you start from home and the time you start work. In (9), for example, the time relation is explicitly expressed in the dependent clause *when she had me*, which provides the limiting point at which you either have or do not have sufficient time to do something.

In addition, sufficiency is often discussed as a modal notion, and several types of constructions have been discussed in this respect, most notably quantifiers like *enough* and *too* and similar expressions (see Malis 2014, for a brief overview). There is some debate in the literature with respect to what type of modality is involved – necessity or possibility. By way of illustration, it is possible, as shown by Beck & Rullman (1999: 261), cited in von Fintel & Iatridou (2007), to paraphrase clauses with the adjective *sufficient* in one of two ways:

(12)   Four eggs are sufficient (to bake this cake).
        a. It is not necessary (given the rules for your cake baking) that you have more than four eggs.
        b. It is possible (given the rules for your cake baking) that you have only four eggs.

Since similar paraphrases are possible with *hinna*, as shown in (13), we feel justified in assuming that *hinna*, too, encodes modal meaning.

(13)   Jane hinner   läsa tre   artiklar på en timme.
       Jane HINNA-pres read three articles in an hour.
       a. It is not necessary (in order to finish three articles) that Jane has
       more than one hour.
       b. It is possible (in order to finish three articles) that Jane has only
       one hour.

Specifically, we follow Beck & Rullman (1999) in assuming that sufficiency is a subtype of possibility. In section 7, we see how this analysis fits with the modality framework proposed by van der Auwera and Plungian (1998), in which sufficiency can be understood as a particular type of participant-external possibility, i.e. possibility in relation to a specified or understood time limitation as opposed to participant-internal possibility, which reflects a person's inherent ability. For now, it suffices to conclude that the correspondences with *have time* mirror a core semantic notion of time sufficiency in *hinna*, which we understand as a participant-external modal notion expressing degree in relation to how much time is required for an event to be possible. In order for the event to be possible, however, it is of course also a requirement that someone is able to do something given sufficient time. In the next section, we consider translations which pick up the semantic notion of ability.

3.2 Ability (as participant-internal possibility)

Table 1 shows that the most common congruent correspondences of *hinna* conveying a clear ability meaning are *can* and *be able to*. An interesting question, however, is whether the ability modals reflect pure ability in these correspondences, or whether ability meaning is conditioned by particular contexts. With this in mind, consider (14) and (15), which illustrate a typical context type for correspondences with ability modals:

(14)   Innan jag <u>hinner</u> eller vill svara fortsätter han: […] (MS1)
       Before I <u>can</u> reply, or even begin to want to, he continues, "Trust
       me." (MS1T)

(15)   Sedan hade jag lämnat dem stående på däck och skyndat mig
        därifrån innan Alva <u>hunnit</u> komma på vad hon skulle svara.
        (JMY1)
        Then I left them standing on the deck and hurried away from there
        before Alva <u>could</u> think of an answer. (JMY1T)

In (14) and (15) *hinna/can* occurs in temporal *before*-clauses, which
are normally interpreted as veradical/factive (see Condoravdi 2010 for a
review of the semantics of such clauses.), and these *before*-clauses
licence negative polarity items like *anything*, thus behaving similarly to
negative clauses. In (14) and (15), the negative meaning is that there is
no answer given. Interestingly, *can* is favored both as translation and
source for *hinna* in these context; 5 of the 10 translations with *can* occur
in such *before*-clauses (cf. examples (4) and (15) above), and 6 of the 13
sources. What these examples illustrate, then, is a kind of negative
actuality conditioned by the temporal clauses. In other words, 'pure'
ability is not negated but rather the ability of replying or thinking of an
answer as conditioned by the *before*-clause.

We hypothesize that the reason for choosing a translation with *can*
rather than *have time* is often the fact that time sufficiency is made
explicit by means of a time limitation and that the correspondences
instead mirror a *presupposed* ability meaning in *hinna*.[4] Our main reason
for regarding ability as a presupposition and not an entailment is that the
presupposed ability meaning persists under negation. Sentence pair
(16ab) illustrates this relation:

(16)   a. Tim hann/hann inte svara.
         Tim HINNA-past/ HINNA-past NEG answer
         'Tim had time to/did not have time to answer.'

        b. Tim kunde svara.
         'Tim was able to answer.'

To specify, the presupposition that Tim is able to answer persists
even if (16a) is negated. It is precisely this presupposed ability that

---

[4] Although *have time* also appears in this context – as illustrated, for instance, by
example (4).

seems to be captured in the cases with negative actuality. In (14), for example, the negative actualization of the act of replying does not mean that the subject referent is unable to act, just that he/she did not manage to do it within the time limitation provided.

Beyond the examples with negative actuality, *can* and *be able to* reflect non-actualized ability for *hinna*, particulary in non-past tense examples. In (17), for example, *be able to* is used for non-actualized ability in a context where the ability can not be expressed by a modal auxiliary in the translation due to the defectiveness of the inflectional paradigm for English modals.

(17) Vi upplever den tekniska utvecklingen som enorm, läkarvetenskapen som allt mer omöjlig att besegra, variationsrikedomen i varuutbudet som allt mer oöverskådligt och avstånden i världen som allt kortare. Vi kan t ex se vilken del av världen som helst inom några sekunder på en TV-skärm i vårt vardagsrum, vi kan gå i ett varuhus under en hel dag utan att <u>hinna</u> se allt och vi byter arbetsplats oftare än någonsin. (BB1)
And all of us normally experience the technical development as enormous, the medical sciences as ever more invincible, the richness of variety of goods and services as more and more difficult to have an overview of, and the distances in the world are ever shorter. We may, for instance, see any part of the world within a few seconds on our television screen in our living-rooms, *we can enter a supermarket and spend one entire day without <u>being able to</u> see everything*, and we change jobs more often than ever before. (BB1T)

The text in (17) is all about the effects of technological development, the availibility of goods and a shrinking world, and it is against this backdrop that the *hinna*-clause is understood. What the highlighted clause conveys is something like 'we lack the ability to see everything that is available in a supermarket since there is simply too much of it' i.e. non-actualized ability.

We conclude that the correspondences with ability seem to reflect a presupposed ability meaning which becomes relevant in contexts where there is a time sufficiency limitation made explicit. Further, many of the correspondences with ability modals involve actuality, but that actuality

does not seem to be a necessary component, particularly not in the non-past examples. We discuss the notion of actuality in the next section.

3.3 Actuality

The material include 17 correspondences with *manage,* all as translations of *hinna*. Curiosly there are no sources with *manage*, which may be an effect of translation bias – *manage* has a relatively clear correspondence in the verb *lyckas* ('succeed'). In line with Karttunen (1971), we regard *manage* as a verb entailing *actuality* meaning. More specifically, Karttunen (1971, 2012) argues that *manage* is a two-way implicative verb, which means that in positive contexts, it is entailed that the event in the complement clause took place, and in negative contexts that it did not take place, as exemplified in (18ab) (adapted from from Baglini and Francez 2015). This differs from so-called factual verbs, such as *admitted*, exemplified in (19), where the truth of the complement clause does not change with the polarity of the complement-taking verb.

(18)   Two-way implicative verb:
       a. Solomon managed to build the temple – Solomon built the temple.
       b. Solomon didn't manage to build the temple – Solomon did not build the temple.

(19)   Factive verb:
       Solomon admitted that he built the temple – Solomon built the temple.
       Solomon did not admit that he built the temple – Solomon built the temple.

   Our data are thus compatible with an analysis of *manage* as an implicative verb. This means that correspondences with *manage* for *hinna* entails the actuality or non-actuality of the event in the complement clause. However, *manage* differs from *hinna* since *actuality* is not necessarily entailed in *hinna* – only in *manage*. That is, although there is potential, and also likelihood, for actuality readings in *hinna*, we argue in the discussion of the correspondences with *can* and *be able to* in 3.2 above that this actuality is not obligatory. Nevertheless, where

manage is used in the translations, it serves as an indication that the actuality implicature of *hinna* goes through in that particular context. Support for this is found in van der Auwera and Plungian (1998: 104) who define *manage* as "participant-internal actuality". To illustrate, even when it takes an explicitly actualized complement like *succeed in*, *manage (to)* seems to express actualized ability, as in the following Googled example, (20):[5]

(20) This was mainly because even if *a family* managed to succeed in clearing the bureaucracy of obtaining the title, their   inability   to fulfill land tax duties meant that they could be dispossessed by the government for nonpayment /.../ (Abdulei et al 2015)

A possible problem with (20) is whether *manage to* in fact gives rise to non-actualized meaning on account of the generic reading of the Subject, *a family*. To explain, as noted by Bhatt (1999) for *was able to*, generic readings of indefinite NPs like *a family* (in 1) favor non-actualised interpretations. In example (l), however, we would argue that the most natural reading of *managed to* is still one of actualized ability. Further, in positive contexts, the correspondences with *manage* also attribute to a participant the *ability* to do whatever the complement clause denotes: If someone managed to to something, it follows that they also had the ability to do it. Understood *ability* is less clear in negative polarity contexts, however, where it can be questioned whether the examples reflect general inability to do something, or a more 'temporal' inability conditioned by the context of situation. We return to this issue below.

In addition to its actuality entailment, *manage to* is typically claimed to presuppose that the event requires some effort on the part of the subject (e.g. Karttunen 1971, Bhatt 1999: 177-179, Aijmer 2004). This observation is supported in our material. In example (21), for instance, *effort* seems presupposed:

(21)   Jag hann lära mig en del... Hon lärde mig en del. (GT1)
        "I managed to learn a bit myself... She taught me a bit." (GT1T)

---

[5] We include a Googled example since the corpus data did not contain clear examples of explicitly actualized complement.

In fact, it seems that this presupposition of effort is central to choosing *manage* as a translation. These translations have in common that there is some effort or uniqueness involved in someone's ability to do something (or not do something) within the time limitation set up. Very often, the examples are from negative polarity contexts. This is illustrated in (22) where the discourse context suggests that the conductor has particular (unique) abilities beyond what we normally expect from conductors:

(22) By the time the driver has come out with his full range of rude words and actually given Johnny a box on the ears without Johnny realizing it, the conductor <u>has managed to</u> get the trolley poles back in place so the gentlemen can get back into their bus and return to their timetable, and allow all the other traffic that has piled up in the meantime to get going again, as the number fifty-two coming from the other direction has also stopped, purely out of sympathy, nothing dramatic. (PP1T)

We conclude that *manage* reflects actualized ability meanings presupposing a notion of effort. The actuality observation is important in light of our understanding of *hinna* as a modal verb: on the one hand, in a definition of (non-epistemic) modality as a means of referring to events/states beyond the here and now, e.g. to the possibility of some event taking place, actualized ability is not, strictly speaking, a modal meaning at all. On the other, the occurrence of so-called actuality entailments (Bhatt 1999) has been associated with modal verbs crosslinguistically, e.g. deontic (semi-) modal auxiliaries such as French *devoir* 'must/have to' (see Hacquard 2014 for discussion), and this suggests that the modality meaning cannot be dismissed purely on the grounds of actuality meaning. In section 7, we address the relation between actualized ability and modal meanings from the point of view of van der Auwera and Plungian's semantic map of modality.

Summing up, what the actuality translations tell us is that *hinna* can, indeed, imply that an event actually occurred, but it is not the case that all instances of *hinna* do this. In other words, *hinna* does not entail actuality, but it typically conveys it. Since we assume that *hinna*'s core meaning is that of time sufficiency, we suggest that actuality is a case of implicature, and that the translations with *manage* pick up this

implicature. It is not absolutely clear whether the implied actuality is tied to the lexical item *hinna*, however, or if it is more generally associated with expressions of sufficiency. We tentatively assume that actuality is a conversational implicature associated with at least some expressions of ability and sufficiency. As an example of a potential expression, we note that the Swedish verb *orka* ('have enough energy to be able to') appears to share the implication of actuality with *hinna*, as evidenced, for example, by a high proportion of English translations with *manage* for *orka* in the ESPC. In addition, Karttunen (1971) has noted that *too* and *enough* may, but need not, give rise to actuality readings. As regards the nature of the exact relation between sufficiency meanings and implications of actuality, we await further research.

4 Double translations and monosemy
Our initial classification in section 2 of the congruent correspondences as reflecting ability, time sufficiency or actuality meanings might suggest that *hinna* is polysemous between these readings, and that the correspondences reflect this. However, we have several reasons to think that the situation is different. For one, we have noted that in some of the negative polarity contexts, *hinna* corresponds to an ability modal which seems to capture a presupposition of ability in *hinna* which is conditioned by a context that makes a time-sufficiency meaning explicit (e.g. temporal dependent clauses), and the implicative verb *manage* seems to reflect a conversational implicature of actualization (and a presupposition of effort) in *hinna*, but since *manage* is an implicative verb, *ability* is also understood. There is thus reason to look beyond the congruent correspondences for clues about how the meaning of *hinna* is mirrored.

An eye-opener in this respect is that, although there were actually very few examples in our data, translators may opt to translate *both* the ability/actuality component *and* the time sufficiency component of *hinna* in separate constructions. Consider (23):

(23)   Jag ser ronden komma men <u>hinner</u> inte varna henne. (PCJ1)
        I see them coming but don't <u>manage</u> to warn her <u>in time.</u> (PCJ1T)

In (23), with an implicit time limitation 'before the round arrived', *hinna* is translated with both *manage to*, which entails actuality and ability, and *in time*, which has no source in the original other than *hinna*. In our opinion, there is nothing even remotely tautologous about the translation, it merely separates the meaning components of *hinna* into distinct expressions. Further, the translation underlines that the *actualized ability* reading is conditioned by limitations, which reflects Teleman et al's insightful description of *hinna* referred to in the introduction, as marking the ability of the subject referent to realize a particular action within a relevant time frame (Teleman et al 1999(4): 297).

The translation in (23) indicates to us that *hinna* is not polysemous, but rather that the three meaning components (time sufficiency, actuality, and ability) are present simultaneously. This means that a serious question arises with respect to the relative ease with which correspondences of *hinna* can be categorized as denoting ability, time sufficiency or actuality, as in Table 1 above. In section 5, we re-examine the data to see how the components can be identified in the correspondences, both in the cases which were initially given a single classification as time sufficiency, ability and actuality and in those translations initially classified as 'zero'.

5 Contextual time limitations

Table 1, was mainly concerned with congruent correspondences associated with *hinna* (in both directions of translation SO→ET and ST←EO), resulting in a relatively neat classification into ability, time sufficiency and actuality correspondences.[6] However, given the nature of time sufficiency, this approach did not really do justice to the correspondences. Nor, in fact, did we analyze the originals in enough detail. In particular, we did not analyze all originals with respect to the presence or absence of an explicit limiting time frame. In going through the correspondences, however, we note that a 'temporal' limiting point is commonly present in the context of the originals, both in the Swedish and in the English examples. Consider, for example, (24) and (25):

---

[6] In addition, we noted accomplishment readings, but these were dismissed from the discussion on the grounds that they may represent a separate meaning of *hinna*.

(24)   Men just den <u>här veckan</u> hade han inte <u>hunnit.</u> (HM2)
       Her husband Robert usually baked all the bread they needed, but
       he <u>hadn't managed to</u> <u>that week</u>. (HM2T)

(25)   He smells of peppermint LifeSavers; underneath that is his
       ordinary smell, of cedarwood lead pencils and wet sand.
       Sometimes he throws up into paper bags, or beside the road if my
       father <u>can</u> stop the car <u>in time</u>. (ST1)
       Det luktar LifeSavers med mintsmak om honom; under det finns
       hans vanliga lukt, av blyertspennor i cederträ och våt sand. Ibland
       kräks han i papperspåsar, eller vid vägkanten om far <u>hinner</u> stanna
       bilen <u>i tid</u>. (ST1)

Both the original and the translation in (24) and (25) contain an explicit
time limitation, *that week* in (24) and *in time* in (25). Initially, the
correspondences were classified as actuality (*manage* in 27) and ability
(*can* in 28), respectively, but clearly, this misses the point that the
actuality/ability is seen in relation to a given time period. The pattern that
emerges now is that the temporal contextual frame is present for
correspondences with both time sufficiency verbs and (actualized) ability
verbs. Thus, we can expect that even the zero correspondences occur
with such contextual time limitations.

    More specifically, upon close examination, it is clear that in roughly
two thirds (20/27) of the zero translations and in three fourths of the zero
sources (29/34), the duration of an event is related to another time period
in one of three ways:

     I.   A specifically defined time period (the relevant time frame
         is made specific)
     II.   A point in time or the beginning of another event limits the
         time period.
     III.   Other event as 'norm' (the relevant time-frame is set by the
         pace of another event)

    Examples (26)-(28) are illustrations of I-III, respectively (the
contextual time limitation is indicated in bold).

(26) Han står tyst **ett kort ögonblick**. **Tillräckligt länge** för att jag ska
     <u>hinna</u> se att någonting skiftar i hans ansikte. (MS1)
     He remains silent **for a moment**, **long enough** for me to see that
     something changes in his face. (MS1T)

(27) Ny kung blev Kristoffer av Bayern som bara <u>hann</u> styra några år
     innan han dog och Karl Knutsson Bonde valdes till kung. (HL1)
     The next king, Christopher of Bavaria, reigned for only a few
     years **before he died** and he was succeeded by Karl Knutsson
     Bonde, who (...). (HL1T)

(28) Hittills hade han inte varit utsatt för några olyckstillbud heller,
     medan Valfrid i sina trasiga och urväxta svenskkängor alltid fick
     springa med huggande mjälte för att <u>hinna</u> över med budkärran
     innan lokomotiven krossade honom som en lus på en tapet.
     (KE2)
     So far they had done their job, he had n't even had any close calls,
     while Valfrid, in his worn-down, outgrown work boots always had
     to rush across the tracks with his wheelbarrow, his heart pounding,
     **to keep the locomotive from crushing him like a louse on the
     wallpaper.** (KE2T)

Moreover, in three of the 'zero' translations (3/27), and of three of the
'zero' sources (3/34), a contextual element not present in the Swedish
version appears in the English version to underline the temporal
limitation. Examples (29) to (30) illustrate addition in such presence (in
bold):

(29) Och förmodligen skulle de inte ens ha <u>hunnit</u> få något larm, de
     skulle inte ens vara där. (JG1)
     Anyway, they would n't have received the alarm **yet**, would n't
     even be there waiting. (JG1T)

(30) To Aristotle **by now** the painting of which he and Homer were part
     was much more than an imitation. (JH1)
     För Aristoteles hade målningen som han och Homeros ingick
     i <u>hunnit</u> bli vida mer än en imitation. (JH1T)

In (29)-(30), the modal meaning of time sufficiency encoded in the auxiliary *hinna* in the Swedish originals is captured in the adverb *yet* in (29), and the prepositional phrase *by now* in (30). These translations can be described as 'divergent' translations (Johansson 2007: 25), i.e. a translation where the form of the original is changed, but meaning is kept, which means that they should be removed from the category zero translation in Table 1. Translations (29) and (30) clearly reflect Time limitation II: a point in time or the beginning of another event limits the time period.

In four of the zero translations (4/27) and five of the zero sources (5/34), we find that there is no clear evidence of a temporal contextual limitation in the sentence. One example is given in (31):

(31)   Jag <u>hann</u> se hennes vita tänder när hon log mot barnet. (LH1)
        (I HINNA.past see her white teeth when she smiled to the child.)
        I saw her teeth **flash** as she smiled at him. (LH1T)

Without further contexts, the sentence in (31) can be interpreted so that *hinna* captures time sufficiency in the sense that the duration of her smile gave the speaker ('I') sufficient time to see 'her teeth'. In addition, it is possible that *hinna* comments on the brevity of the smile (i.e., the smile was so quick the speaker only barely had enough time to see her teeth). Some support for this analysis is the addition of the verb *flash* in the English translation. As pointed out to us by an anonymous reviewer, however, an equally likely explanation is that the time limitation is present in the wider context of the discourse. In this example, the previous sentence gives a hint of a limitation in that it explains how the woman returns inside her hut and is thus no longer visible to the narrator.

6. A note on time sufficiency and ability
The point of departure for this study was the observation that *hinna* conveys two modal meanings: *time sufficiency,* a specialized type of participant-external possibility and *ability*, a sub-category of participant-internal possibility. The results show that apart from some examples where *get to* translates the structure [*hinna* + place adverb], time-sufficiency is nearly always present in translations of *hinna,* either by a congruent translation correspondence such as *have time* or by contextual

time limitations that are carried over from the original or made more explicit by means of a divergent translation correspondence such as *by then*. This result is important, since it suggests that, except in constructions with [*hinna* + place adverb], *hinna* is not polysemous between ability and time sufficiency readings. Instead, we have suggested that time sufficiency is part of the core meaning and that *hinna* additionally presupposes ability and implies actuality via a conversational implicature.

Our proposal that time sufficiency, rather than ability, represents the core meaning of *hinna* is supported in cases where the ability presupposition does not go through. In our material, we have identified one context which clearly blocks an ability reading: with inanimate subjects, ability cannot be part of the meaning, i.e., it is not presupposed, and instead we get a pure time sufficiency reading. Also, in (32), the event is actualized, which additionally shows that ability and actuality are distinct.

(32)   I skaftet veckade sig skinnet och Franz hade legat lungsjuk så länge att det <u>hunnit</u> komma damm i vecken. (KE2)
       The uppers were creased, and Franz had been ill for so long that dust had accumulated in the crease. (KE2T)

In (32), *hunnit* seems to function as a comment on the length of the time period, but since we already have the adverbial *so long*, it adds very little in terms of meaning compared to the alternative without the auxiliary, which is the alternative chosen in the translation. In our material, we have only four examples with an inanimate Subject, and they have in common that they appear to add some minor comment, possibly of an aspectual nature, on the length of a period that has some relevance in the context. In addition, although our data include no such examples, *hinna* can occur with 'dummy' subjects like weather-*it*, as in (33):

(33)   Det hann börja regna innan utflykten var slut.
       It HINNA-past start to-rain before the-picknick was over.

Thus, with regard to the meaning of *hinna*, the English correspondences in this material suggest that Swedish *hinna* is a monosemous implicative verb where the modal category time sufficiency

implies the modal category ability in contexts with agentive subjects. The implication is asymmetric – whereas enough time implies ability, insufficient time does not imply inability. In contexts with inanimate subjects, on the other hand, the syntactic context may block the ability reading. In addition, in constructions with [*hinna* + adverb], often translated with *get to,* we may have a different, non-modal meaning of *hinna*, which suggests that *hinna* can be polysemous. We have also noted that evidence for regarding ability and time sufficiency as modal categories can be considered problematic in examples which involve actualization.

7. Hinna*'s semantic map*

In this section, we argue that the observations about *hinna* as an implicative modal verb make sense viewed through the lens of van der Auwera and Plungian's (1998) proposal for a universal concepual space for modality: Modality's Semantic Map. We first introduce the notion of semantic map, and then place *hinna* on the map.

A semantic map represents a method for comparing languages with respect to meaning/use differences, and is based on empirical observations in a number of languages. As pointed out by Haspelmath (2003), one of the advantages of semantic maps is that they do not "imply a commitment to a particular choice among monosemic and polysemic analyses". What semantic maps do instead is that they show how expressions can be represented as covering a geometric semantic space, and thus they provide a means for comparison of form-meaning relations across languages. In the words of Boye (2010: 9): "The structure of a semantic map is an empirical result obtained by first identifying and generalizing over a number of comparable meanings across languages and then studying which meanings are directly related to each other in terms of synchronic polyfunctionality or diachronic change". Importantly, this means that entailments and presuppositions may be represented in separate categories on the map – as noted by van Leusen (2012: 426): what serves as an implication of lexical items at one point in time can become "conventionalised and part of their lexical description'" in the course of time.

We can now take a look at van der Auwera and Plungian's semantic map for modality (van der Auwera and Plungian 1998: 80). On this map,

modality is defined in terms of the notions of *possibility* and *necessity* whose meanings/uses form a continous space on the map. The map is interesting since it can be used to visualize the semantic boundaries of *hinna* in the universal conceptual space of modality, i.e., as a visualization beyond notions of polysemy and monosemy, or, for that matter, entailments, presuppositions and implicatures. To specify, according to van der Auwera and Plungian (1998), expressions for modality have moved or developed across adjacent categories on the map diachronically, and/or they can be polysemous or monosemous between two adjacent meanings synchronically. Figure 1 illustrates continuously related meaning categories in the semantic space of possibility, which is the space of interest for *hinna* (based on van der Auwera and Plungian 1998). The boxes with dotted pattern illustrate the semantic space coverd by *hinna*.

The semantic boundaries suggested in Figure 1 are based on our discussion of *hinna* as mirrored through its English correspondences. Two observations are of particular interest. Firstly, as suggested in the correspondences with the construction [verb + space adverbial], such as *arrive* and *get there*, *hinna* can be used for premodal meanings such as *arrive at* and *finish*. The map thus supports our suggestion in section 2 that an accomplishment meaning of *hinna* corresponding to a [verb + place adverbial] construction in English preceeds the other more modal meaning/uses of *hinna*. Secondly, regarding the correspondences with *have time*, *manage*, *be able to* or *can*, we have argued that *hinna*, by virtue of conversational implicature, covers the pre-modal adjacent meanings "participant-internal actuality" and "participant-external actuality", but also, by virtue of presupposed meaning, the modal meanings participant-internal possibility and, by entailment, participant-external possibility.
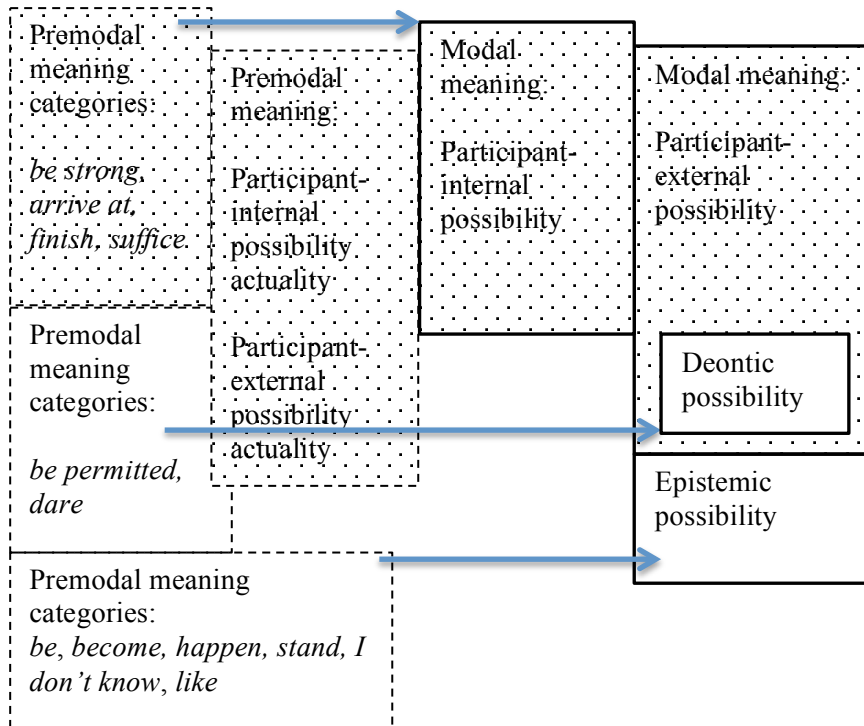
Figure 1 *Hinna* on Possibility's Semantic Map (based on van der Auwera and Plungian's Modality's Semantic Map (1998)). *Hinna* covers the patterned area (dots). Dotted lines represent premodal meanings and solid lines modal meanings. Arrows show how premodal lexical categories can grammaticalize into modal meanings.

Figure 1, then, provides a solution to the 'actualization' problem noted particularly for past-tense examples of *hinna*. In the map, participant-internal possibility actuality and participant-external possibility actuality are regarded as pre-modal categories related to the modal categories in terms of adjacency. We note further that *hinna* typically combines participant-internal possibility (ability) and participant-external possibility (time sufficiency), and argue that the relation between these meanings is one of implication, i.e. time sufficiency presupposes ability. This suggests to us that *hinna* has ventured into the domains of participant-internal and participant-external possibility in 'one go'. That is, since time-sufficiency meaning is

obligatorily present for *hinna*, we find it unlikely that *hinna* has had, or can have, a purely participant-internal possibility meaning. The pre-modal category mirrored in the translations with *arrive* and *get there*, on the other hand, might represent a separate meaning of *hinna*. From the perspective of the semantic map, there is nothing contradictory in *hinna* behaving both as a polysemous verb (separate meanings) and a monosemous verb (implication/vagueness). As pointed out by Haspelmath (2003), for instance, an expression can be polysemous with respect to some distinctions on a map and vague (monosemous) between others.

9 Conclusion

By studying the correspondences of the Swedish verb *hinna,* we believe we have managed to show the usefulness of translation corpora both in generating or confirming hypotheses about the semantic properties of lexical items, and in separating different approaches to phenomena like multiple meanings. By approaching the corpus correspondences from two different angles; focusing first on congruent correspondences, and then on semantic features in the context of such correspondences, we were able to demonstrate that *hinna* is best understood as encoding two meaning components, i.e. time sufficiency as entailment and ability as presupposition, but also typically conversationally implying actuality.

Figures 2 and 3 provide a schematic conclusion of parts of our analysis. Figure 1 illustrates how the meaning components of *hinna* are captured in the correspondences with ability verbs, such as as *can* or *be able to,* and Figure 3 illurates how the same components are captured in the correspondences with *manage*:
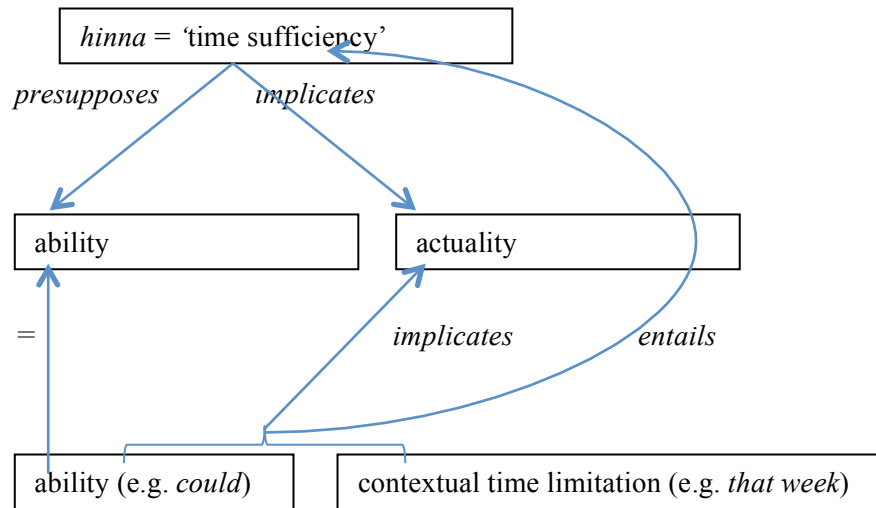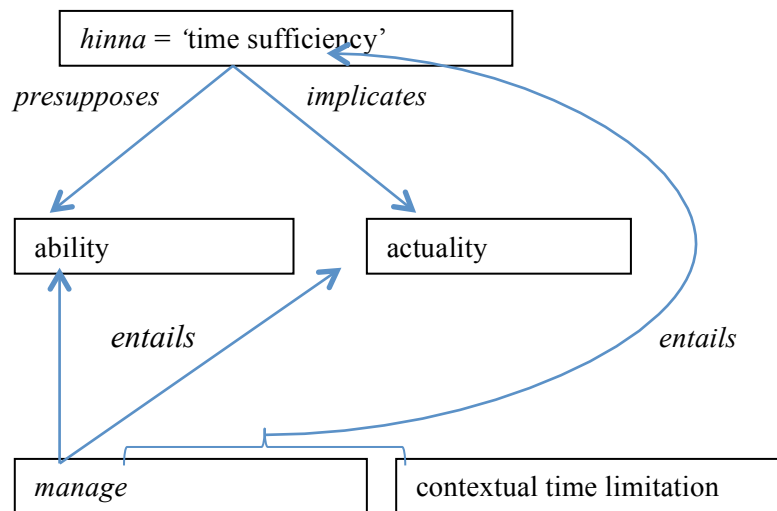
Figure 2 *Hinna* corresponding to ability verbs

Figure 3 *Hinna* corresponding to *manage*

From top to bottom Figures 2 and 3 illustrate our analysis of 'time sufficiency' as the core semantic component of *hinna*, and also how *hinna* presupposes ability and conversationally implicates actualization. From bottom to top they show how English correspondences (both translations and sources) can take different paths to these meaning components. Figure 2 shows how ability translations such as *can* together with a contextual time limitation entail time sufficiency and conversationally imply actuality, and Figure 3 shows how *manage* entails actuality on its own (as does *was able to*, if Bhatt 1999 is right), but can only imply time sufficiency in case the context supplies an expression denoting a time limitation such as *in time*. Correspondences directly expressing time sufficiency, such as *have time*, are not illustrated in Figures 2 and 3, but here we assume that the meaning time sufficiency is accessed directly.

We believe that our description of the meaning components of *hinna* is compatible with an analysis of *hinna* as an "implicative" verb (Karttunen 1971, 2012), even if we do not claim that actuality is an entailment of *hinna*, but rather a conversational implicature. Further, since we claim that *hinna* is unambiguously a sufficiency modal, enriched by implications, we assume that its core semantics may be derived via a Kratzerian analysis (in which underspecification and contextual determination of the core semantics is used to deal with the alleged ambiguity of modals like *must* or *can*). However, the exact relation between our approach and the formal treatments of 'polyfunctional' modality proposed by Kratzer (1981, 1991) remains unclear.

We argue, though, that time sufficiency and ability are contiguous modal meanings, in the sense that they occupy neighbouring areas in a semantic map of modality. Concretely, from the point of view of van der Auwera and Plungian's Modality's Semantic Map (1998), we understand both to be non-epistemic possibility senses; ability is participant-internal possibility, whereas time sufficiency is participant-external possibility. We also claim that actualized ability, although strictly speaking pre-modal, is contiguous with 'true' modal ability. For our present purposes, the distinction between pre-modal and true modal senses is not crucial, however. More important is the observation that the different senses are contiguous, which is compatible with our discussion of how the time sufficiency and ability/actuality meanings are related.

An important advantage of viewing meaning relations in terms of semantic maps is that it illustrates that even implied and presupposed meanings can be analysed as contiguous. This is interesting from a contrastive/typological perspective since it offers a possibility to illustrate, as suggested by van Leusen (2012), how implications can become part of the core meaning of words by way of conventionalization or specialization in one language, but not another, resulting in a situation where corresponding verbs in two closely related languages differ only in 'the strength of their implicative inference'. We take studies that zoom in on such contrastive differences to be a fruitful area of future exploration. Another area for further research that we have only touched upon in this paper is how entailments, presuppositions and implications of verbs interact with tense and aspect differences.

References

Abdulai, Raymond Talinbe, Franklin Obeng-Odoom, Edward Ochieng, and Vida Maliene. Eds. 2015. *Real Estate, Construction and Economic Development in Emerging Market Economies*. London: Routledge.

Aijmer, Karin. 2004. "The semantic path from modality to aspect: *'be able to'* in a cross-linguistic perspective." *Corpus Approaches to Grammaticalization in English*. Eds. Hans Lindqvist and Christian Mair. Amsterdam/Philadelphia: John Benjamins. 57-79.

Altenberg, Bengt and Karin Aijmer. 2000. "The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies." *Corpus Linguistics and Linguistic Theory*. Eds. Christian Mair and Marianne Hundt. Amsterdam and Atlanta, GA: Rodopi. 15-33.

Baglini, Rebekah and Itama Francez. 2015. "The implications of managing." *Journal of Semantics*. doi: 10.1093/jos/ffv007.

Beck, Sigrid and Hotze Rullmann. 1999. "A flexible approach to exhaustivity in questions." *Natural Language Semantics* 7: 249-298.

Bhatt, Rajesh. 1999. *Covert Modality in Non-Finite Contexts*. Ph.D. Thesis, University of Pennsylvania.

Boye, Kasper 2010. "Semantic maps and the identification of cross-linguistic generic categories: Evidentiality and its relation to epistemic modality." *Linguistic Discovery* 8(1): 4-22.

Chesterman, Andrew. 1998. *Contrastive Functional Analysis*. Amsterdam: John Benjamins.

Condoravdi, Cleo. 2010. "NPI licensing in temporal clauses." *Natural Language and Linguistic Theory* 28: 877-910.

Hacquard, Valentine. 2014. "Actuality entailments." Under review for *Companion to Semantics*. Eds. L. Matthewson, C. Meier, H. Rullmann, T. E. Zimmermann. London: Wiley.

Haspelmath, Martin. 2003. "The geometry of grammatical meaning: semantic maps and cross-linguistic comparison." *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure*. Ed. Michael Tomasello. Mahwah NJ: Lawrence Erlbaum. 211-242.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora*. Amsterdam/Philadelphia: John Benjamins.

Karttunen, Lauri. 1971. "Implicative verbs." *Language* 47(2): 340-358.

Karttunen, Lauri. 2012. "Simple and phrasal implicatives." *Proceedings of the First Joint Conference on Lexical and Computational Semantics* (*SEM), ACL: Montreal. 124-131.

Kratzer, Angelika. 1981. "The notional category of modality," *Words, worlds, and contexts: new approaches in word semantics*. Eds. H.-J. Eikmeyer and H. Rieser. Research in text theory 6. Berlin: de Gruyter, 38-74.

Kratzer, Angelika. 1991. "Modality." *Semantics, An International Handbook of Contemporary Research*. Eds. A. von Stechow and D. Wunderlich, Berlin: de Gruyter. 639-650.

Teleman, Ulf, Staffan Hellberg and Erik Andersson. 1999. *Svenska Akademiens grammatik, vol. 1–4*. Stockholm: Svenska Akademien.

van der Auwera, Johan and Vladimir A. Plungian. 1998. "Modality's semantic map." *Linguistic Typology* 1998(2): 79-124.

von Fintel, Kai, and Sabine Iatridou. 2007. "Anatomy of a modal construction." *Linguistic Inquiry* 38: 445-83.

van Leusen, Noor. 2012. "The accommodation potential of implicative verbs." *Logic, Language and Meaning. 18th Amsterdam Colloquium, The Netherlands, December 19-21, 2011, Revised Selected Papers*. Eds. M Aloni, V. Kimmelman, F. Roelofsen, G. Sassoon, K Schulz & M Westera. Berlin-Heidelberg: Springer, 421-430.