

Per Studia Contrastiva ad Astra: Looking Back, Leaping Forward

Workshop convenors:

Lot Brems (University of Liège, lbrems@uliege.be)
Lobke Ghesquière (University of Mons & KU Leuven, lobke.ghesquiere@umons.ac.be)
Gudrun Vanderbauwhede (University of Mons, gudrun.vanderbauwhede@umons.ac.be)

Using contrastive corpora for investigating speech acts from an English-Swedish perspective - the case of requests

Karin Aijmer
(University of Gothenburg, karin.ajmer@sprak.gu.se)

A major issue in pragmatics is to identify the linguistic manifestations of speech acts such as requesting, offering, suggesting, etc. It is also a topic to which corpora and corpus linguistics can contribute. Corpus linguists have been successful in investigating the functions of speech acts having a conventionalized form (cf. e.g. Deutschmann 2003 on apologies). However, other methodologies are needed to identify all (or a representative sample of) the patterns associated with a speech act function. A promising trend in monolingual corpus-based speech act studies has been to proceed from the definition of a speech act function to identify how the speech act is realized by manually reading through the text or by utilizing the results from previous research (see e.g. Pöldvere et al. 2022). The aim of my presentation is to extend the function-to-form methodology to the contrastive investigation of speech acts. The case study is an investigation of how polite requests are performed in English and Swedish using the English-Swedish Parallel Corpus as data for the analysis (Altenberg and Aijmer 2000).

Requests have been classified by Searle (1976) as directives having certain properties or felicity conditions which must be fulfilled by the patterns categorized in this way. The hearer must, for example be able and willing to carry out an action. It would be difficult to search for all the requestive forms manually in order to investigate their correspondences in the other language. The cross-cultural common core can therefore be taken to be interrogatives or declaratives with a modal auxiliary (and usually a second person subject). I therefore began to search for the occurrences of these patterns in the corpus in the English and Swedish original texts. The English search items are, for instance, *can you*, *could you*, *will you*, *you could*, *you would*, etc. but also *please*, *maybe*, *I think*, *just* which can be supposed to collocate with a polite request in their mitigating functions. I also searched for patterns with *please* (or one of the other markers) manually excluding examples where it was not followed by a request. Specifically, the Swedish correspondences of utterances containing *please* as a politeness marker contained patterns associated with requesting (which might be unexpected since they differed from the patterns in the English originals) such as *jag måste be er* ('I must ask you'), *var vänlig (snäll, bussig, god) och* ('be kind and'), *ni kan väl* ('you can I suppose'), *det går bra att* (lit. 'it goes well that'), *du kanske skulle* ('you perhaps should'). The preliminary findings suggest that we can get a rich description of the realizations of the speech act of requesting in the compared languages

highlighting similarities and differences between the languages by using the corpus both as a comparable corpus and as a translation corpus.

References

- Altenberg, B. and K. Aijmer. 2000. 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 Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20)*. Freiburg im Breisgau 1999. Amsterdam/Atlanta: Rodopi. 15-33.
- Deutschmann, M. 2003..*Apologising in British English*. Institutionen för moderna språk, University of Umeå.
- Pöldvere, N., Felice, R. D., & Paradis, C. 2022. *Advice in conversation: Corpus pragmatics meets mixed methods*. Cambridge University Press.
- Searle, J. 1976. A classification of illocutionary acts. *Language in Society* 5 (1): 1-23.

Issues of comparability and sameness in contrastive critical discourse studies

Niall Curry

(Manchester Metropolitan University, n.curry@mmu.ac.uk)

Contemporary corpus-based contrastive linguistics is epitomised by concerns with comparability, data quality, and sameness. Theoretical notions of the *tertium comparationis*, equivalence, and sameness and identity govern how we validate comparisons in contrastive studies (Krzyszowski, 1984), questions of methodological directionality and convergent and divergent approaches influence research outcomes (Chesterman, 2007), and issues of alignment in data design, data quality, and representation (Johansson, 2003) pose challenges for corpus linguists engaging in contrastive work. Over the last 30 years, there has been an increased effort to merge contrastive and corpus methodologies (e.g., Hasselgård, 2020; Johansson, 2003; McEnery & Xiao, 2010) and unpack approaches to contrastive analysis in a range of subfields of corpus linguistics (e.g., corpus-assisted discourse studies, Vessey, 2013). However, in the wider literature, theoretical constructs in contrastive linguistics have had a relatively limited impact on multilingual and comparative research, while the use of corpus linguistics approaches appears to be growing across and beyond applied linguistics (Pérez-Paredes & Curry, 2024). In the context of critical discourse studies, for example, the use of corpus linguistics is largely normalised. Yet, in contrastive and multilingual critical discourse studies, the role of notions such as equivalence, sameness, and the *tertium comparationis* are largely unaddressed. As such, there is a need to critically assess the relevance of these fundamental concepts in critical discourse studies and delineate their effective operationalisation therein.

To operationalise these key concepts, in this talk I discuss two empirical studies, focusing on a contrastive analysis of 1) Brexit and 2) COVID-19 discourses in expert communication in English, French, and Spanish. For each analysis, I use themed corpora of academic news blog posts from *The Conversation*'s English, French, and Spanish language sites. *The Conversation* is an international site used by academics to disseminate their research to the public. With over 40 million monthly readers, of whom over 80% are non-academics, these texts are largely designed to inform the public of research developments across the spectrum of academic disciplines. These blog posts have been found to demonstrate cultural differences across languages and reconstruct cultural and ideological perspectives (e.g., Curry, 2024). As such, they offer a valuable site for understanding the nature of expert communications, globally. For the analysis of Brexit discourses, I use a comparable corpus to conduct a convergent critical discourse analysis based on a comparison of keywords in each corpus. For the analysis of COVID-19 discourses, I use a parallel corpus of academic news blogs posts and their translations to conduct a divergent analysis, using transitivity analysis. Through these studies, I return to foundational concepts in contrastive linguistics to bolster corpus approaches to critical discourse studies. In so doing, I draw attention to the affordances of contrastive research and its theoretical underpinnings for offering a complex and layered understanding of globalised and localised social challenges.

References

- Chesterman, A. (2007). Similarity Analysis and the Translation Profile. *Belgian Journal of Linguistics*, 21(1), 53-66.
- Curry, N. (2024). Questioning the climate crisis: A contrastive analysis of parascientific discourses. *Nordic Journal of English Studies*, 23(2), 235-267.
- Hasselgård, H. (2020). Corpus-based contrastive studies: Beginnings, developments and directions. *Languages in Contrast*, 20(2), 184-208.
- Johansson, S. (2003). Contrastive linguistics and corpora. In Granger, S., Lerot, J., & Petch-Tyson, S. (Eds.). *Corpus-based approaches to contrastive linguistics and translation studies* (pp. 31-44). Brill.
- Krzeszowski, T. P. (1984). Tertium comparationis. *Contrastive linguistics. Prospects and problems*, 301-12.
- McEnery, T., & Xiao, R. (2010). *Corpus-based contrastive studies of English and Chinese*. Routledge.
- Pérez-Paredes, P., & Curry, N. (2024). Epistemologies of corpus linguistics across disciplines. *Research Methods in Applied Linguistics*, 3(3), 100141.
- Vessey, R. (2013). Challenges in cross-linguistic corpus-assisted discourse studies. *Corpora*, 8(1), 1-26.

From *star* (N) to *star* (V): A contrastive study of cognate noun/verb pairs in English and Norwegian

Jarle Ebeling

(University of Oslo, jarle.ebeling@usit.uio.no)

Signe Oksefjell Ebeling

(University of Oslo, s.o.ebeling@ilos.uio.no)

Previous cross-linguistic studies of specific cognate word pairs in English and Norwegian have shown varying degrees of overlap in frequency and use across the languages, e.g. Ebeling (2017) on *bring* and *bringe* and Ebeling (2024) on *see* and *se*. This paper widens the scope and investigates a set of cognate English and Norwegian heterosemous words that, through conversion, “express related meanings across multiple word classes” (Shao et al. 2023: 321), e.g. *star* (N) and *star* (V). In a diachronic study of heterosemy in recent English (1920s-2010s), Shao et al. (2023) compiled a list of 877 heterosemous nouns and verbs from the Corpus of Historical American English. This list will serve as the starting point of the present English-Norwegian contrastive analysis, with a focus on pairs that have etymologically related counterparts in Norwegian.

In contrast to English, conversion in Norwegian often requires some affixation, e.g. *hat* (N) vs. *hate* (V) ‘hate’, although zero-derivational pairs also exist, e.g. *pumpe* (N) and *pumpe* (V) ‘pump’. Thus, conversion by zero-derivation was not a requirement in the Norwegian material. This is in line with Huddleston & Pullum’s (2002) understanding of conversion.

After identifying corresponding cognate noun/verb pairs, we aim to establish how and to what extent such pairs are used in English and Norwegian. We do this by searching for each pair in the English-Norwegian Parallel Corpus+, a bidirectional corpus that will enable us to establish the items’ Mutual Correspondence (Altenberg 1999). In this investigation of more than 100 noun and verb pairs we will be in a position to provide a more comprehensive and accurate account of how cognates behave cross-linguistically. To our knowledge, a study of cognate heterosemous noun/verb pairs has not been done on this scale before, as most previous studies have focused on individual cognate pairs, often belonging to one word class only.

More specifically, we seek answers to the following research questions:

1. To what extent do noun and verb uses of the cognates correspond to each other in translation?
2. What factors tend to trigger a higher or lower degree of Mutual Correspondence (MC)?

Preliminary results show that the nouns generally tend to have a higher MC than the verbs, indicating that the verbs more commonly develop diverging polysemies and/or conditions of use. It is expected that these findings will be substantiated in the analysis of the complete set of noun/verb pairs.

Although similar trends are reported at the overall level, cognate noun/verb pairs show contrastive differences at the individual level in terms of frequency and MC. The final part of the study, addressing the second research question, will therefore consist of case studies analysing individual pairs that show different degrees of MC in the hope of revealing features

that may contribute to this variation in MC between two closely related languages. While MC is meant as an initial step in establishing cross-linguistic equivalence, a more detailed cross-linguistic analysis of the kind proposed here is necessary to shed further light on precisely what is common to the languages compared and what sets them apart.

References

- Altenberg, B. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In H. Hasselgård and S. Oksefjell (eds), *Out of Corpora: Studies in Honour of Stig Johansson*, Amsterdam: Rodopi, 249–268.
- Ebeling, S.O. 2017. Bringing home the bacon! A contrastive study of the cognates *bring/bringe* in English and Norwegian. *Kalbotyra* 700: 104–126.
- Ebeling, S.O. 2024. Seeing through Languages and Registers: A Closer Look at the Cognates *See* and *Se*. In Cermakova, A., H. Hasselgård, M. Malá and D. Šebestová, Denisa (eds), *Contrastive Corpus Linguistics. Patterns in Lexicogrammar and Discourse*. London: [Bloomsbury Academic](#). 29–61.
- Huddleston, R. and G.K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Shao, B., J. Zheng and H. De Smet. 2023. The blurring of the boundaries: changes in verb/noun heterosemy in Recent English. *Corpus Linguistics and Linguistic Theory* 20(2), 321–346.

Verb insertion in translations as a sign of grammaticalisation in progress

Thomas Egan
(University of Inland Norway, thomas.egan@inn.no)

When analysing translations of expressions in parallel corpora, one normally distinguishes between syntactically congruent and non-congruent translations. There may also be expressions in the original text which the translator has omitted altogether, or expressions in a translation lacking a correspondence in the original. These are both referred to as ‘zero translations’ (see Johansson 2007: 26). It is the second type of zero translation that is the topic of this presentation. An example of such a translation is (1), taken from the English–Norwegian Parallel Corpus (ENPC), in which the ingressive aspect signalled by *begynne* (begin) in the translation is not present in the English original.

(1) Inside his head the synapses were shutting down. (DF1)

I hodet *begynte* synapsene å slukne. (DF1T)

‘In the head began the synapses to go out.’

When an expression is undergoing grammaticalisation, it normally undergoes semantic bleaching (see, for example, Heine and Narrog (2010: 406)). A textbook example of a grammaticalised expression in English is the *going to* future, as in (2).

(2) “Jeg *skal* møte henne imorgen.” (LSC2)

“I shall meet her tomorrow.”

“I’m *going to* see her tomorrow.” (LSC2T)

In (2) the expression *going to* does not translate a verb of inherently directed motion (Levin 1993: 263), but the present tense form of the Norwegian modal verb *skulle*, here coding a planned action (Faarlund et al. 1997: 604). The translation is syntactically congruent, since in both the source and target text the first verb licences an infinitive clause.

(1) and (2) are, respectively, clear-cut examples of a zero translation and a congruent translation of a grammaticalised expression. In this presentation I examine less clear-cut cases, investigating a handful of constructions which may be in the process of grammaticalising. These include English constructions containing the matrix verbs *help* (Mair 1995) and *fail* (Mackenzie 2008, Egan 2016) and Norwegian pseudo-coordinate constructions headed by the posture verbs *stå* (stand) and *sitte* (sit) (Kinn et al. 2018, among many others). All of these constructions have already been subject to some contrastive analysis, the *fail* construction in Egan (2018), the *help* construction in Egan (2024), and the Norwegian constructions in Tonne

(1999), Johansson (2009) and Ebeling (2015). Here I revisit the constructions, on the basis of data from one and the same corpus, the ENPC, and consider all examples in the translations in which the first verb, be it matrix verb or first coordinate, lacks a lexical correspondence in the original text. The research questions is as follows:

Can the insertion of a second verb in the translation of a one-verb construction be taken as an indication of the grammaticalisation of (that verb in) the resulting two-verb construction?

Preliminary analysis shows that the corpus contains enough evidence on which to base the discussion of this question.

References

- Ebeling, S. O. 2015. A contrastive study of Norwegian pseudo-coordination and two English posture-verb constructions. In *Cross-Linguistic Perspectives on Verb Constructions*, S. O. Ebeling and H. Hasselgård (eds), 29–57. Newcastle: Cambridge Scholars Publishing.
- Egan, T. 2016. The subjective and intersubjective uses of FAIL TO and ‘not fail to’. In *Aspects of Grammaticalization: (Inter)subjectification and Pathways of Change*, H. Cuyckens, L. Ghesquière and D. van Olmen (eds), 168–196. Berlin: De Gruyter.
- Egan, T. 2018. The FAIL TO Construction: A Contrastive Perspective. *Bergen Language and Linguistics Studies* 9 (1). <https://doi.org/10.15845/bells.v9i1.1525>.
- Egan, T. 2024. Simple and complex help constructions in English and Norwegian: a contrastive study. *Languages in Contrast*. 24 (1), 84 – 108.
- Faarlund, J. T., Lie S. and Vannebo K. I. 1997. *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.
- Heine, B. and H. Narrog. 2010. Grammaticalization and Linguistic Analysis. In *The Oxford Handbook of Linguistic Analysis*, B. Heine and H. Narrog (eds), 401–423. Oxford: Oxford University Press.
- Johansson, S. 2007. *Seeing Through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.
- Johansson, S. 2009. Norwegian-pseudo-coordination with verbs of posture in translation into English and German. In *From Will to Well: Studies in Linguistics offered to Anne-Marie Simon-Vandenberghe*, S. Slembrouck, M. Taverniers and M. Van Herreweghe (eds), 279–291. Ghent: Academia Press.
- Kinn, T., K. Blenselius and P. Andersson. 2018. Posture, location, and activity in Mainland Scandinavian pseudocoordinations. *Cognitextes*. <https://doi.org/10.4000/cognitextes.1158>
- Levin, B. 1993. *English verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Mackenzie, J. L. 2008. Failing without trying. *Jezikoslovlje* 9(1-2): 53–85.

- Mair, C. 1995. Changing patterns of complementation, and concomitant grammaticalisation, of the verb help in present-day English. In *The Verb in Contemporary English: Theory and Description*, B. Aarts and C. F. Meyer (eds), 258–72. Cambridge: Cambridge University Press.
- Tonne, I. 1999. A Norwegian progressive Marker and the Level of Grammaticalization. *Languages in Contrast* 2 (1), 131–159.

Verbal expressions of habituality in English and Norwegian: forms and correspondences

Hilde Hasselgård
(University of Oslo, hilde.hasselgard@ilos.uio.no)

Habitual expressions “describe a situation which is characteristic of an extended period of time, so extended in fact that the situation referred to is viewed not as an incidental property of the moment but, precisely, as a characteristic feature of a whole period” (Comrie 1976: 27-28). While there is no fully grammaticalized habitual aspect in verb systems of English and Norwegian, both languages have periphrastic expressions of habitual meaning as defined above. Examples are *used to*, *tend to* (*I used to help him out / This tends to work well*) and *pleie*, *bruke* (*Faren pleide å gi ham en bok... – His father used to give him a book...*). Besides catenative expressions, modal auxiliaries (e.g. English *would*, Norwegian *kunne*) and complement clause constructions can express habituality, e.g. Norwegian *det hender at* ‘it happens that’ and English *be known to*. Previous comparisons of English and Norwegian have focused on a single expression of habituality and its correspondences, e.g. Bjerga (1998), who studied Norwegian *pleie* and its English correspondences, and Johansson’s (2005) examination of *det hender at* ‘it happens that’. Lund (2007) investigated whether *used to* could always be translated by *pleie*. These studies found that habitual expressions are rarely translated congruently between English and Norwegian. Similarly, Altenberg (2007) found low mutual correspondence between English *used to* and Swedish *bruka*.

The present study considers verbal expressions of habituality in both Norwegian and English and their correspondences. The material comes from the English-Norwegian Parallel Corpus, in which the following expressions have been searched for in original texts: the habitual catenatives *pleie* and *bruke* + infinitive and the English expressions *used to*, *tend to*, *be known to* and *would*. The aims are the following: to compare the use and distribution of habituality expressions within and between the languages; to survey the range of translation correspondences of each expression; and to identify potential factors influencing the choice of correspondence.

Preliminary results show that *pleie* far outnumbers *bruke* in Norwegian originals. In English *would* is most frequent, followed by *used to*, *tend to* and *be known to*. In sum, the English habituais are more frequent than the Norwegian ones. Recurrent noncongruent correspondences include adverbials, e.g. *usually*, *sometimes*, *alltid* (‘always’) *før* (‘before’). This aligns with Altenberg’s (2007) results, and was also apparent in Hasselgård (2007), in which usuality adverbials sometimes corresponded to verbal expressions. Furthermore, translations into both languages frequently omit the habituality marker and use a simple tense instead, and Norwegian may use the perfect aspect to mark anteriority. Altenberg (2007) shows that co-occurring time adverbials and the tense and dynamicity of the verb phrase influence the choice of translation between English and Swedish. These factors will also be examined in the proposed study, which in addition aspires to build on and connect the findings of the previous studies and to identify a possible division of labour between verbal, adverbial and other expressions of habituality in both languages examined.

References

- Altenberg, Bengt. 2007. Expressing past habit in English and Swedish. In *Functional Perspectives on Grammar and Discourse: In Honour of Angela Downing*, ed. by C.S. Butler, R.H. Downing, J. Lavid, 97-128. Amsterdam: John Benjamins.
- Bjerga, Trude Davidsen. 1998. Continuative and habitual aspect in English and Norwegian. With special reference to the English verb *keep* and the Norwegian verb *pleie*. Unpublished MA thesis, University of Oslo.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- 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(S1). <https://doi.org/10.35360/njes.8>
- Johansson, Stig. 2005. Some aspects of usuality in English and Norwegian. In *Semiotics from the North: Nordic Approaches to Systemic Functional Linguistics*, ed. by E. Maagerø and K.-L. Berge, 69-86. Oslo: Novus.
- Lund, Karianne. 2007. Kan 'used to' alltid oversettes med 'pleide'? : en sammenlikning av used to og pleie som uttrykk for habitualitet, med fokus på oversettelse fra engelsk til norsk. MA thesis, University of Oslo. <http://urn.nb.no/URN:NBN:no-16964>.

Corpus

The English-Norwegian Parallel Corpus, see
<https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/>

Shifts from noun phrase postmodification to premodification in academic writing: Towards conditions and contexts of change

Marie-Pauline Krielke
(Saarland University, mariepauline.krielke@uni-saarland.de)

Isabell Landwehr
(Saarland University, isabell.landwehr@uni-saarland.de)

Relative clauses (RCs) provide explicit noun phrase (NP) post-modification and over time, they have become less favored in scientific English. Facilitated by greater reliance on background knowledge, they have been replaced by denser, less explicit NP modifications like attributive adjectives and compounds (Hundt *et al.*, 2012; Biber & Gray, 2016). In German, RCs initially increase before declining later than in English, reflecting German's delayed establishment as a primary scientific language (Krielke, 2021). However, RCs persist in both languages. This study investigates when RCs are replaced and when they survive in scientific discourse, focusing on RCs with copula verbs and predicative adjectives (e.g., “*the element which is solid*,” RC+A). Their development is compared to attributive adjective + noun constructions (e.g., “*the solid element*,” A+N). An information-theoretic approach guides this analysis, based on the following assumptions:

- a) A+N usage increases in both English and German, while RC+A declines in English but initially rises and later declines in German.
- b) A+N usage becomes increasingly predictable (lower surprisal) compared to RC+A in both languages.
- c) Surviving RC+A constructions become more predictable over time, occurring in entrenched (Bybee, 2002) contexts where grammatical necessity prevents more condensed expressions.

Our data set for English is the Royal Society Corpus, (RSC Version 6.0 Open, Fischer *et al.*, 2020) consisting of texts from the Philosophical Transactions and Proceedings of the Royal Society between 1650 – 1920. The German data set is the scientific portion of Deutsches Textarchiv (DTAW, Geyken *et al.*, 2018) comprising scientific books covering 1600 – 1890. Both datasets are annotated with parts-of-speech and 4-gram surprisal (Shannon, 1948). The information-theoretic notion of surprisal captures a word's predictability given its context (here: the three preceding words). High surprisal indicates low predictability and vice versa.

First results show that A+N constructions increase steeply in both languages, while RC+A constructions are less frequent. In English, RC+A constructions decline, whereas in German, they initially rise and then decrease after 1740.

Surprisal of RC+A adjectives is constantly higher compared to A+N in both languages, likely due to their lower frequency. Despite declining frequency in English, RC+A surprisal decreases, indicating increasingly entrenched usage: Surviving RC+A constructions feature entrenched forms with further specifications (e.g., prepositional phrases or infinitival complements), while simple “*x, which is y*” constructions become rare. In German, surprisal

increases for both constructions, suggesting less entrenched adjective usage. Although adjectives directly followed by “*sein*” decrease, the variability (entropy, Shannon, 1948) of other continuations remains stable. These findings support the hypothesis that, in expert scientific communication, such simple constructions are avoided to reduce redundancy. Their partial survival is motivated by their inability to be prepended, unlike standalone adjectives.

Future analyses will explore whether RC+A constructions in highly specialized contexts (e.g., “*the plane which is perpendicular*”) serve as precursors to A+N constructions and examine the role of surprisal in driving their shift toward denser encodings. Specifically, we will assess whether decreasing surprisal in RC+A constructions facilitates the emergence of corresponding A+N forms.

References

- Biber, D., and Gray, B. 2016. Grammatical complexity in academic English: Linguistic change in writing. Cambridge University Press.
- Bybee, J. 2002. Main clauses are innovative, subordinate clauses are conservative: consequences for the nature of constructions. In Joan Bybee & Michael Noonan (eds.), *Complex sentences in grammar and discourse: essays in honor of Sandra A. Thompson*, 1-18. Amsterdam/Philadelphia: John Benjamins.
- Fischer, S., Knappen, J., Menzel, K., and Teich, E. 2020. The Royal Society Corpus 6.0: Providing 300+ Years of Scientific Writing for Humanistic Study. *Proceedings of the 12th Language Resources and Evaluation Conference*, 794–802.
- Geyken, A., Boenig, M., Haaf, S., Jurish, B., Thomas, C. and Wiegand, F. 2018. Das Deutsche Textarchiv als Forschungsplattform historische Daten in CLARIN. In *Digitale Infrastrukturen die germanistische Forschung* (= Germanistische Sprachwissenschaft um 2020, Bd. 6), H. Lobin, R. Schneider and A. Witt (eds.), 219–248. Berlin/Boston: De Gruyter.
- Hundt, M., Denison, D., and Schneider, G. 2012. Relative complexity in scientific discourse. *English Language and Linguistics*, 16(2), 209–240.
- Krielke, M.-P. 2021. Relativizers as markers of grammatical complexity: A diachronic, cross-register study of English and German. *Bergen Language and Linguistics Studies*, 11(1), 91–120.
- Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. *The Bell System Technical Journal*.

Swedish compound nouns and English noun sequences – a perfect match?

Magnus Levin

(Linnaeus University, magnus.levin@lnu.se)

Jenny Ström Herold

(Linnaeus University, jenny.strom.herold@lnu.se)

Vasiliki Simaki

(Lund University, vasiliki.simaki@englund.lu.se)

In a previous cross-linguistic study (Ström Herold & Levin, forthcoming), we investigated English noun sequences, i.e. juxtaposed nouns (*world war*; *health care reform initiative*), with their German and Swedish correspondences, the results indicating that around 70% of the correspondences are (solid) noun-noun compounds (*Weltkrieg/världskrig*), irrespective of language and translation direction. Thus, there appears to be a fairly strong cross-linguistic correlation between these two types of structures – at least with this methodological approach. What we do not know at present is if this observed correlation is an interference effect from English. To address this knowledge gap, we reverse the study, instead taking Swedish compounds as the point of departure.

Therefore, this study aims to explore: (i) the proportion of Swedish compound nouns in relation to their English correspondences (noun sequences or compounds), (ii) the distributions of other types of correspondences such as postmodifying prepositional phrases and premodifying adjectives, and (iii) what these results tell us about language-specific preferences and translation effects.

In this study, an extended tagged version of the Linnaeus University English-German-Swedish corpus (LEGS) non-fiction corpus is used (Ström Herold & Levin 2019; forthcoming). LEGS consists of, e.g., popular science, biographies and self-help books amounting to approximately half a million words of each source language. Using a Python script, we retrieve words tagged as nouns (and their translations), which are then classified, manually removing all non-compound nouns.

We will be considering the same variables as in our previous study: compound noun length (two-part, three-part etc.), common vs. proper nouns as first elements (e.g., *law degree/juristexamen* vs. *Yale degree/Yaleexamen*) and the semantic relations holding between the parts of the nouns (e.g., time relation, purpose relation; cf. Teleman et al. 1999: II: 44–45 for Swedish, and Biber et al. (2021 [1999]: 582) for English). As for the first variable, a corpus study by Carlsson (2004: 75), contrasting Swedish and German newspaper language, showed that two-part compounds constitute more than 90% of her material.

Our preliminary findings suggest that Swedish compounds are rendered as English noun sequences (*familjemedlemmar* [‘family-members’] > *family members*) or solid compounds (*grundvatten* [‘ground-water’] > *groundwater*) in proportions similar to those of the previous study. As for ‘non-compound’ correspondences, English appears to use slightly more premodifying adjectives as correspondences to Swedish compounds (*flingsalt* [‘flake-salt’] >

flaked salt) than in our previous study. Swedish compound nouns seem to be shorter than English noun sequences and less frequently contain proper nouns as first elements, as compared to findings in Ström Herold & Levin (2019; forthcoming). Finally, ‘simple’ English noun correspondences (*vetemjöl* [‘wheat-flour’] > *flour*) are rare (cf. the high frequencies identified by Nessel (2018) in Norwegian-Russian contrast).

Our study will deepen the state of knowledge regarding the similarities and differences between Swedish and English noun phrase structures, an area that is still under-researched for this language pair. It will also try to disentangle the effects of source-language norms or restrictions and translation-induced changes.

References

- Biber, D., S. Johansson, G. Leech, S. Conrad, & E. Finegan. 2021 [1999]. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Carlsson, Maria. 2004. *Deutsch und Schwedisch im Kontrast: Zur Distribution nominaler und verbaler Ausdrucksweise in Zeitungstexten* [Göteborgs germanistiska Forskningen 43]. Göteborg: Acta Universitatis Gothoburgensis.
- Nessel, Tore. 2018. When a single word is enough: Norwegian compounds and their Russian counterparts. *Slovo. Journal of Slavic Languages, Literatures and Cultures* 59, 61–72.
- Ström Herold, Jenny & Levin, Magnus. 2019. *The Obama presidency, the Macintosh keyboard and the Norway fiasco: English proper noun modifiers in German and Swedish contrast*. *English Language and Linguistics*, 23 (4), 827–854.
- Ström Herold, Jenny & Levin, Magnus. Forthcoming. *Climate change and Harvard students: English noun sequences and their German and Swedish correspondences*. To appear in *Corpora* 18 (3).
- Teleman, Ulf, Andersson, Erik & Hellberg, Staffan. 1999. *Svenska Akademiens Grammatik*. Volume II. Stockholm: Norstedts.

Complexity of the noun phrase in English and Czech children's literature

Markéta Malá

(Charles University, malamaff4@gmail.com)

While children's literature is often considered a simplified version of fiction for adults (Thompson & Sealey 2007), there appear to be relatively few studies actually measuring grammatical complexity of fiction for children (e.g., Puurtinen 1998, Montag 2019, Dawson 2023). Complexity measures have been used to analyse genre and register differences (Biber et al. 2011, 2024), the development of children's writing and reading proficiency (Hsiao et al. 2023, 2024), and especially to assess L2 students' proficiency in English (e.g. Bulté & Housen 2012).

This work-in-progress study sets out to explore some of the factors which contribute to syntactic complexity in children's literature in two typologically distinct languages, English and Czech. I focus on the structure of noun phrases headed by nouns, which appears to mirror the developmental stages in language proficiency, with heavy postmodification (especially by prepositional phrases and nonfinite clauses) being indicative of high levels of complexity (Biber et al. 2011).

The research draws on data from the English and Czech sections of the parallel corpus InterCorp. Its latest edition has been annotated using the universal dependencies framework (Rosen 2023) and several measures of syntactic complexity (at the level of the sentence and text) and lexical diversity. The replacement of language-specific morphological tagging by language-uniform morpho-syntactic annotation now makes it possible to expand the scope of contrastive studies to include direct comparison of syntactic structure and complexity across languages. The study is based on two small comparable sub-corpora (0.5 million tokens each) of InterCorp comprising English and Czech original fiction for children.

In both languages, noun phrases in children's fiction were found to be generally shorter, with fewer layers of embedding than in fiction for adults. At the same time, noun phrases in English children's fiction are, on average, longer than those in Czech children's books (3.2 and 3.35 words per phrase, respectively), with similar maximum depth of embedding. This suggests the impact of the analytical character of English as opposed to predominantly synthetic Czech: the category of definiteness typically remains unexpressed in Czech, and the relations between the head and a postmodifying noun may be expressed by case suffixes rather than by prepositions.

As shown by Biber et al. (2024), however, apart from structural distinctions, syntactic functional distinctions have to be considered when studying complexity. The preliminary results indicate that a higher proportion of noun-headed phrases function as the subject in Czech than in English. The fact that the subject can occupy the clause-final position in Czech may then contribute to heavier postmodification within the Czech subject noun phrases. The two languages were also found to differ in the forms of postmodification. The higher proportion of non-finite clauses in English, as opposed to finite relative clauses, can be interpreted in relation to the preference of Czech for finite verb predicates, which make it possible to express the verbal grammatical categories on the lexical verb (cf. Dušková 2015).

References

- Biber, D., Gray, B. & K. Poonpon (2011) Should We Use Characteristics of Conversation to Measure Grammatical Complexity In L2 Writing Development?, *TESOL Quarterly* 45:1, 5-35.
- Biber, D., Larsson, T. & G. R. Hancock (2024) The linguistic organization of grammatical text complexity: comparing the empirical adequacy of theory-based models, *Corpus Linguistics and Linguistic Theory* 20:2, 347-373.
- Bulté, B. & A. Housen (2012) Defining and Operationalising L2 Complexity, in *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA*, ed. A. Housen, F. Kuiken, & I. Vedder, Amsterdam/Philadelphia: John Benjamins.
- Corpus InterCorp*, version 16ud from 17 September 2024. Ústav Českého národního korpusu FF UK, Praha. Available from <http://www.korpus.cz>
- Dawson, N. J., Hsiao, Y., Tan, AWM, Banerji, N. & K. Nation (2023) Effects of Target Age and Genre on Morphological Complexity in Children's Reading Material, *Scientific Studies of Reading*, 27:6, 529-556.
- Dušková, L. (2015) *From Syntax to Text: the Janus Face of Functional Sentence Perspective*. Praha: Karolinum.
- Hsiao, Y., Dawson, N. J., Banerji, N. & K. Nation (2023) The nature and frequency of relative clauses in the language children hear and the language children read: A developmental cross-corpus analysis of English complex grammar, *Journal of Child Language* 50:3, 555-580.
- Hsiao, Y., Dawson, N. J., Banerji, N. & K. Nation (2024) A corpus-based developmental investigation of linguistic complexity in children's writing, *Applied Corpus Linguistics*, 4:1, 1-14.
- Montag, J. L. (2019). Differences in sentence complexity in the text of children's picture books and child-directed speech. *First Language*, 39:5, 527-546.
- Puurtinen, T. (1998). Syntax, Readability and Ideology in Children's Literature. *Meta* 43:4, 524-533.
- Rosen, A. (2023) The InterCorp parallel corpus with a uniform annotation for all languages. *Jazykovedný časopis*, 74:1, 254-265.
- Thompson, P. & A. Sealey (2007) Through Children's Eyes? Corpus Evidence of the Features of Children's Literature, *International Journal of Corpus Linguistics* 12:1, 1-23.

Academic voices past and present: first-person pronouns in English and Norwegian research articles

Sylvi Rørvik

(University of Inland Norway, sylvi.rorvik@inn.no)

This study examines the frequency and rhetorical functions of first-person pronouns in English and Norwegian research articles (RAs). Numerous studies have investigated the use of first-person pronouns as indications of authorial presence in academic texts in a range of languages, including English, Spanish, Swedish, and Lithuanian, as well as many others (cf. e.g. Hyland 2001, Sheldon 2009, McGrath 2016, Hyland & Jiang 2017, Šinkūnienė 2018, Carrió-Pastor 2020, Wheeler et al 2021, Ädel 2022, and Dixon 2022). However, Norwegian academic prose has not been extensively studied. One seminal exception is Fløttum et al's work from 2006, which examined the use of various features including first-person pronouns in Norwegian, French, and English RAs (dating from the early 1990s to the early 2000s) from three fields: economics, linguistics, and medicine. Fløttum et al identified disciplinary differences in the use of first-person pronouns.

Using Fløttum et al's study as the starting point, the present study investigates English and Norwegian RAs from the fields of linguistics and education, including 50 RAs in each field in each language, i.e. 200 in total. These articles have all been peer-reviewed and accepted for publication, and are therefore considered to be representative of English and Norwegian academic prose, even if the native-speaker status of the authors has not been ascertained (cf. Carrió-Pastor 2020: 19).

The aim of the study is partially to enable a diachronic comparison with Fløttum et al's results by employing more recent linguistics material (dating from 2015-2024), but the inclusion of RAs from the field of education serves a primarily pedagogical aim: to provide an empirical basis for teaching and supervising students in the field of education who are required to write academic papers in both English and Norwegian, which is the case for Norwegian students who aim to become teachers of English. The research questions are as follows:

1. To what extent are there cross-linguistic and cross-disciplinary differences in the *frequency* of first-person subject pronouns in English and Norwegian RAs in the fields of linguistics and education?
2. To what extent are there cross-linguistic and cross-disciplinary differences in the *rhetorical functions* of first-person subject pronouns in English and Norwegian RAs in the fields of linguistics and education?

Preliminary results regarding the frequency of first-person pronouns indicate that the subcorpora are characterized by within-group variation, but that both language and discipline seem to play a role in the use of first-person pronouns: In the English material, first-person plural subject pronouns are markedly more frequent in linguistics than in education. In Norwegian the disciplines are very similar, but the frequency for education is twice as high as that observed for English, and for linguistics it is approximately 1/3 higher than that found in the English material. The presentation will also include an overview of the rhetorical functions

expressed by the first-person subject pronouns in the material. This analysis builds on the four roles outlined by Fløttum et al (2006: 81, 83-84): writer, researcher, arguer, and evaluation.

References

- Ädel, A. 2022. "Writer and reader visibility in humanities research articles: Variation across language, regional variety and discipline." *English for Specific Purposes* 65, 49-62.
- Carrió-Pastor, M.L. 2020. "Variations in the Use of Self-Mentions in Different Specific Fields of Knowledge in Academic English." In M.L. Carrió-Pastor (ed.), *Corpus Analysis in Different Genres: Academic Discourse and Learner Corpora*. Taylor & Francis Group, 13-32.
- Dixon, T. 2022. "Proscribed informality features in published research: A corpus analysis." *English for Specific Purposes* 65, 63-78.
- Fløttum, K., T. Dahl, & T. Kinn. 2006. *Academic Voices: Across languages and disciplines*. John Benjamins Publishing Company.
- Hyland, K. 2011. "Humble servants of the discipline? Self-mention in research articles." *English for Specific Purposes* 20, 207-226.
- Hyland, K. & F. Jiang. 2017. "Is academic writing becoming more informal?" *English for Specific Purposes* 45, 40-51.
- McGrath, L. 2016. "Self-mentions in anthropology and history research articles: Variation between and within disciplines." *Journal of English for Academic Purposes* 21, 86-98.
- Sheldon, E. 2009. "From one *I* to another: Discursive construction of self-representation in English and Castilian Spanish research articles." *English for Specific Purposes* 28, 251-265.
- Šinkūnienė, J. 2018. "*I* and *we* in Lithuanian, Lithuanian English and British English research writing." In P. Mur-Dueñas (ed.), *Intercultural Perspectives on Research Writing*. John Benjamins Publishing Company, 59-79.
- Wheeler, M.A., E. Vylomova, M.J. McGrath, & N. Haslam. 2021. "More confident, less formal: stylistic changes in academic psychology writing from 1970 to 2016." *Scientometrics* 126, 9603-9612.

Contrastive Analysis of English and German in a Construction Grammar Framework

Peter Uhrig

(University of Erlangen-Nuremberg, peter.uhrig@fau.de)

Thomas Herbst

(University of Erlangen-Nuremberg, thomas.herbst@fau.de)

We attempt to show how a Construction Grammar (CxG) framework (Goldberg 2019; Hoffmann 2022; Herbst & Hoffmann 2024) can be used to carry out analyses of two closely related languages – German and English. We will build on earlier work on contrastive linguistics (e.g. Burgschmidt & Götz 1974; König & Gast 2018) in that we regard it as essential that any contrastive study should be based on descriptions of the two languages that were carried out independently from one another so as not to prioritize one language over the other. In our view, the model of CxG is particularly well suited for contrastive analyses because it is based on constructions as form-meaning pairings, although we would argue that in the field of contrastive analysis it would be wrong to say that the semantic side of grammatical constructions had been neglected in the structuralist work of the 1960s and 1970s. In contrast to at least some of the very early approaches within contrastive linguistics, we would not argue that contrastive analyses would enable one to predict errors L2-learners are likely to make.

The area of description we intend to focus on is that of argument structure constructions in English and German, which was captured in a lot of descriptive work in the form of complementation patterns (Quirk et al. 1985, and also Hunston & Francis 2000) for English and in terms of valency for German (e.g. Helbig & Schenkel 1968 or Schumacher et al 2004). We will address issues such as the following:

To what extent can one identify corresponding, equivalent or parallel constructions in the two languages, if one assumes an approach that sees constructions as language specific? We would argue that we can identify a double object (i.e. ditransitive) construction in German and in English and that there are clear parallels between them in that they have 4 slots: NP (subject) – V – NP (indirect object) – NP (direct object) with comparable roles (AGENT OR AFFECTOR – ACTION – RECIPIENT – THEME/PATIENT/AFFECTED), which allow us to establish a certain degree of correspondence between the German and English ditransitive constructions. However, the identification of the three different NP slots is driven by word order in English and by case in German. We are going to critically discuss Croft's (2022) comparative concepts in the light of this comparison. Another difference between the German and the English ditransitive constructions is revealed by comparing their collo-profiles, i.e. a corpus-based and frequency-related analysis of the verbs that occur in them (cf. www.constructicon.de; Herbst & Hoffmann 2024). On the basis of different English and German argument structure constructions, we will argue that the comparison of such collo-profiles is a key element of a contrastive analysis in this area, which, of course, was impossible to carry out in pre-corpus times, and which reveals important information about the difference between two languages that – combined with empirical error analysis making use of learner corpora, for instance – can be exploited in the design of foreign language teaching materials and textbooks.

Works cited

- Burgschmidt, Ernst & Dieter Götz. 1974. *Kontrastive Linguistik Deutsch/Englisch*. München: Hueber.
- CASA|Con. 2023-. *A Constructicon of the English Language*, ed. by Thomas Herbst, Thomas Hoffmann, Peter Uhrig, Armine Garibyan & Stephanie Evert. www.constructicon.de.
- Croft, William. 2022. *Morphosyntax: Constructions of the World's Languages*. Cambridge: Cambridge University Press.
- Goldberg, Adele E. 2019. *Explain Me This*. Princeton: Princeton University Press.
- Helbig, Gerhard & Wolfgang Schenkel. 1968. *Wörterbuch zur Valenz und Distribution deutscher Verben*. Leipzig: Enzyklopädie.
- Herbst, Thomas & Thomas Hoffmann. 2024. *A Construction Grammar of the English Language. CASA – a constructionist approach to syntactic analysis*. Amsterdam/Philadelphia: Benjamins.
- Hoffmann, Thomas. 2022. *Construction Grammar: The Structure of English*. Cambridge: Cambridge University Press.
- Hunston, Susan & Gill Francis. 2000. *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: Benjamins.
- König, Ekkehard & Volker Gast. 2018. *Understanding English-German Contrasts*. 4th edition. Berlin: Erich Schmidt Verlag.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Schumacher, Helmut, Jacqueline Kubczak, Renate Schmidt & Vera de Ruiter. 2004. *VALBU – Valenzwörterbuch deutscher Verben*. Tübingen: Narr.