

Introduction

Elisabet Engdahl & Robin Cooper

Elisabet Engdahl, Department of Swedish Language, Göteborg University, Box 200, SE-405 30 Göteborg, Sweden. E-mail: engdahl@svenska.gu.se

Robin Cooper, Department of Linguistics, Göteborg University, Box 200, SE-405 30 Göteborg, Sweden. E-mail: cooper@ling.gu.se

This special issue of the *Nordic Journal of Linguistics* is devoted to Comparative Nordic Semantics. Whereas much research has been carried out on comparative syntax, morphology and phonology in the Nordic languages, much less work has been done on the comparative semantics of these languages. But the fact that some of the Nordic languages, namely the Scandinavian ones, Danish, Faroese, Icelandic, Norwegian and Swedish, are historically, lexically and structurally very similar means that they provide an interesting target for semantic research. Are there systematic semantic differences between these languages? If so, are the formal semantic analytic tools that have been developed mainly for English and German sufficiently fine-grained to account for the differences among the Scandinavian languages? These were some of the questions asked in the research project Comparative Semantics for Nordic Languages (NORDSEM), which was funded by the Joint Committee of the Nordic Research Councils for the Humanities in 1998–2001 and which involved researchers at the Copenhagen Business School, Göteborg University and the University of Oslo.¹ Two of the papers in this issue (by Carl Vikner and Kjell Johan Sæbø) derive directly from the NORDSEM project whereas the third paper, by Erich Round, pursues some issues investigated during the project by Joakim Nivre and published in *Nordic Journal of Linguistics* 25:1 (2002).

In his paper on the semantics of Scandinavian ‘when’-clauses, **Carl Vikner** investigates the use of temporal connectives corresponding to English *when* in the Scandinavian languages. Danish makes a systematic distinction between *da*, which is only used in past episodic clauses, and *når*, used in past and present habitual clauses as well as in future clauses. Vikner’s semantic analysis takes *da*-clauses to be similar to definite noun phrase reference, i.e. they presuppose that there is a single eventuality in the discourse situation which is described by the *da*-clause. *Når*-clauses on the other hand are taken to be similar to indefinite noun phrases, introducing variables over eventualities which may be bound by generic operators and hence are appropriate for habitual sentences. In addition to the detailed semantic analysis of Danish *da*- and

når-clauses, Vikner briefly investigates the uses of the cognate connectives *da/då* and *når/när* in Norwegian and Swedish, as they show up in corpora. In Swedish, *när* is used in both episodic and habitual sentences, much like the English *when*. Norwegian seems to be an intermediate case between Danish and Swedish. Whether a Swedish or a Norwegian *når*-clause should be given an episodic or habitual interpretation is determined by the context and knowledge of the world, much in the same way as languages without articles nevertheless distinguish between definite and indefinite interpretations of noun phrases.

Since the three Scandinavian languages investigated, as well as German and Dutch, have two separate words for *when*, some interesting topics for further research arise from Vikner's paper. Have the other Scandinavian languages previously been like Danish, or is the Danish systematic distinction an innovation? Does the apparent ongoing change in Norwegian reveal anything about the way speakers assign temporal interpretations?

Erich Round's paper centers on the use and meaning of Swedish *någon/nån* and English *some*, and thus complements Nivre's 2002 article, in which he used a contrastive Scandinavian perspective as a key to analyzing the system of indefinite determiners in Swedish (Nivre 2002:4). Round shows that English *some* and Swedish *någon* are both used when the identification supplied by the indefinite noun phrase is inadequate in some respect as for instance when the speaker doesn't know the identity of the referent or deems this information to be irrelevant. In addition they can both be used as quantitative hedges but in slightly different ways. Swedish *någon vecka* 'some week' can mean 'approximately one week', which the English *some week* cannot mean. In English, *some* can only combine with expressions denoting more than one, e.g. *some seven days*. In both languages, the use of *some/någon* can convey uncertainty as to whether a particular description holds of the referent, e.g. *some military officer (or something)/någon militär (eller något)*. The main purpose of Round's comparative analysis is to identify which of these meanings are 'coded' in the lexical items, i.e. context-invariant, and which meanings arise through various inferential processes, in particular conversational implicatures, as used by Grice.

Most of Round's data come from two contemporary spoken language corpora consisting of Australian and Swedish teenagers' conversation. In order to establish the direction of development, Round also makes use of Shakespeare's plays, a convenient but maybe not representative historical corpus.

Both Vikner and Round use language corpora as a way of finding relevant data for their semantic analyses. **Kjell Johan Sæbø** argues strongly in his article that it is in fact necessary for semanticists to use corpus data in their work. Taking the history of research on Free Choice items such as *any* in English as an example, he argues that semanticists need to check the predictions made by their hypotheses against corpora. Only if the analysis accounts for all substantial classes of clear cases in a random sample will it be valid. Corpus analysis provides one way of minimizing the

‘nocuous idealizations’ that may arise when the researcher focuses on a limited set of data. Sæbø then goes on to test some hypotheses using Scandinavian Free Choice items which have the form *(h)vilken . . . som helst* ‘which . . . ever’, where *helst* is the superlative of *hellre*, ‘rather’. He notes that the majority of these items occur either in explicit modal contexts or in contexts which implicitly or explicitly are intensional. This finding provides the ground for the unified hypothesis put forward in Sæbø (2001). The point made in the present paper is that corpora can be used not only to falsify previous accounts but also constructively, as a source for general hypotheses and as indicators for what constitutes the core cases as opposed to isolated constructed counter examples.

All three articles in the present issue use corpora of various kinds, and this seems to be symptomatic of a lot of current linguistic research, including semantics. They mainly use monolingual corpora but Sæbø also uses a parallel English-German corpus when he checks the cross-linguistic validity of the so-called Universalist hypotheses. Parallel aligned corpora are excellent tools for lexicographers, syntacticians and translators, but clearly also for semanticists. In this context, it is worth mentioning that a couple of projects are underway to establish multilingual parallel corpora involving Scandinavian languages. A project at the University of Oslo is extending the *English-Norwegian Parallel Corpus* into the *Oslo Multilingual Corpus*² (OMC), which will include original texts in English, Norwegian, German and French, with translations into the other three languages. At Språkbanken (the Bank of Swedish) at Göteborg University, the SALT³ project is establishing a corpus of originals and translations of fiction in Swedish, Dutch, English, French, German, Italian and Russian.

The OMC and SALT initiatives will provide useful tools for contrastive analysis of individual Scandinavian languages and other European languages. However, from the perspective of comparative Scandinavian research, the obvious resource, a parallel Scandinavian corpus, is still missing. Proposals for such a corpus have been made and will hopefully be realized before long. Furthermore, improvement of the automatic alignment methods will facilitate the building of what might be called the next generation of parallel corpora, viz. parallel treebanks. A treebank is a collection of syntactically parsed sentences or utterances. Since it contains structural information, it is often more useful for syntactic and semantic research than plain text corpora. Several projects to build monolingual treebanks are underway; within the *Nordic Treebank Network*,⁴ there are plans to establish a Nordic parallel treebank, which would constitute an extremely useful resource for semanticists and syntacticians alike.⁵

NOTES

1. For more information about the NORDSEM project, the participants and other publications, see <http://www.ling.gu.se/projekt/nordsem/>

2. <http://www.hf.uio.no/german/sprik/english/corpus.shtml>
3. Språkbankens Arkiv för Länkade Texter (the Bank of Swedish Archive of Linked Texts), <http://spraakbanken.gu.se/lb/salt/>
4. A network, funded by the Nordic Language Technology Program, see <http://w3.msi.vxu.se/~nivre/research/nt.html>
5. For an example of what kinds of representations might be available in parallel treebanks, see Volk & Samuelsson (2004).

REFERENCES

- Nivre, Joakim. 2002. Three perspectives on Swedish indefinite determiners. *Nordic Journal of Linguistics* **25.1**, 3–47.
- Sæbø, Kjell Johan. 2001. The semantics of Scandinavian Free Choice items. *Linguistics and Philosophy* **24**, 737–787.
- Volk, Martin & Yvonne Samuelsson. 2004. Bootstrapping parallel treebanks. In *Proceedings of the Workshop on Linguistically Interpreted Corpora (LINC), COLING-2004*, Geneva, 63–69.