

## The Nordic Word Order Database: New methods for studying variation within and between speakers and varieties in North Germanic

The North Germanic languages have undergone massive changes over the last 800 years, visible for instance in the change from OV to VO word order. Furthermore, the Mainland Scandinavian (Msc.) varieties (Swedish, Norwegian and Danish) and their dialects have followed fairly similar patterns in their development, e.g. loss of case, verb agreement and V-to-I movement, whereas the Insular Scandinavian (Isc.) varieties (Icelandic and Faroese) have kept these features (although, more recently, Faroese has moved towards the morphosyntactic patterns of the Msc. varieties) (Holmberg & Platzack 1995, Vikner 1995). In many cases, all MSc. varieties have moved in the same general direction, but at a more detailed level, we find diverging paths, resulting in a fairly large amount of word order variation within the language area. Some of the variation can be described as between-language, but, in fact, most of the variation occurs *within* the different varieties. We do not yet know if certain syntactic phenomena are restricted to certain dialects/registers or reflect optionality/intraspeaker variation. This uncertainty is especially clear when considering today's language situation, where speakers get linguistic input from a wide variety of dialects and registers, and from mass media and written material. Given this, the question we address is: are these (relatively infrequent) word order variants stably associated with dialect groupings, in similar ways to omnipresent linguistic variables such as intonation, segmental phonology and high-frequent inflectional material?

In this talk, we present the Nordic Word Order Database (NWD), the rationale behind it and the methods used in data elicitation and data analysis. NWD is an online database with a user-friendly search interface, hosted by Tekstlab at the University of Oslo to be launched in early 2019. It contains elicited production data from speakers of all the Scandinavian languages, including several different dialects. So far, fieldwork has been conducted in seven different locations, and data from altogether around 250 participants (age: 16-60) have been collected (appr. 50 000 sentences in total). Given its size, and the systematic methods used in the data collected, some core questions about linguistic variation in the modern speech communities can be answered (e.g., is syntactic variation conditioned mainly by "region" or "register", or is the variation mainly within or between individuals).

The database contains information about eight different syntactic macro-phenomena, that have been discussed extensively in the Scandinavian syntactic literature (see e.g. Thráinsson 2001, Svenonius 2002, Wiklund et al. 2007, Westergaard et al. 2017): **(1) Subject shift**: the placement of pronominal and NP subjects with respect to sentence adverbials; **(2) Object shift**: the placement of pronominal and NP objects with respect to sentence adverbials; **(3) Long object shift**: the placement of light pronominal objects with respect to NP subjects; **(4) Particle shift**: the placement of verbal particles with respect to pronominal and NP objects (and subjects); **(5) "Long" particle shift**: the placement of verbal particles with respect to inverted DP subjects; **(6) Embedded V2**: the placement of finite verbs with respect to sentence adverbs in embedded non-factive clauses; **(7) V-to-I movement**: the placement of finite verbs with respect to sentence adverbs in embedded questions; **(8) V3 in non-subject questions**: the placement of the finite verb with respect to the subject in non-subject questions. For all of the phenomena, at least one of the Scandinavian languages has developed a word order pattern that is categorically different from the other languages. There is also a relatively well-established description of the diachronic development of the relevant variables. In addition, the phenomena are known to be infrequent in natural speech. The database, therefore, provides an important addition to corpora – it gives a new means to investigate both inter- and intra-individual variation, contra dialect and language variation.

Two production experiments are used for data collection, and they include largely the same sentences in all languages. Both build on a simple design whereby a participant reads a sentence presented on a screen aloud at a given cue. The participant is then provided with a new start of the sentence and is asked to transform the sentence. Five different transformations are used, in order to avoid priming: (a) inversion: *The baker bought not the book yesterday* → *Yesterday...* (*bought {the baker} not {the baker} a book* (tests phenomena 1, 3 and 5); (b) tense change and inversion (phenomena 1, 2, 3 and 5); (c) passive to active (phenomena 2 and 4); (d) main to embedded clause (phenomenon 7) and (e) embedded to main clause (phenomenon 8). 5-15 items of each phenomena are included in the experiment, which makes it possible trends within and between individuals within the

same dialect. The experiments are recorded, and the sound files are segmented and coded for word order, and will be accessible through NWD.

The preliminary results indicate that both syntactic variation and prosodic/intonational properties elicited in the experiment match the patterns found in spoken language corpora. Moreover, the results reveal considerable variation both within and between speakers, but also surprising stability within the individual languages: syntactic phenomena that show a variable pattern in one language have a fully categorical distribution in a closely related language. However, with respect to dialectal variation, the preliminary results tie only very few of the variables tested to specific dialects: optionality in object shift is found in speakers of all Swedish dialects tested; particle shift (around NP objects) is variable in all Norwegian dialects tested; embedded V2 (in information structure neutral contexts) is variable in speakers from different regions of the Faroe Islands. V3 in non-subject question seems to be a genuine dialect phenomenon, but even within the dialects it is used, it appears to be optional.

As noted, all of the phenomena tested show some variability within a dialect. To some extent, this has been shown in previous corpus studies as well (see e.g. Andréasson 2009 on object shift in Swedish, Westergaard 2011 on subject shift in Norwegian). However, since the phenomena are so infrequent in spontaneous production (for instance, non-sentence initial, non-pronominal subjects in sentences with negation are rare), it is almost impossible to investigate the factors behind the variation between or within individual speakers using corpus methods. In our experimental data, we find from the same dialect area with significantly different patterns of variation. In western Sweden, we can, for instance, find speakers who have almost consistent subject shift, but also speakers that only shift non-pronominal subjects in a very small number of cases. The database allows us to test whether and how the individual patterns cluster with the other phenomena tested. Further, since the database includes high quality sound files, we can also test whether word order variation correlates with variation in prosodic structure and intonation, as has been claimed for both object shift (Josefsson 2010) and particle shift (Larsson & Lundquist 2014).

## References

- Andréasson, Maia. 2009. Pronominal object shift — not just a matter of shifting or not. *Working Papers in Scandinavian Syntax* 84: 1-20.
- Bentzen, Kristine, Anderssen, Merete & Waldmann, Christian. 2013. Object shift in spoken Mainland Scandinavian: A corpus study of Danish, Norwegian and Swedish. *Nordic Journal of Linguistics* 36(2): 115–151.
- Josefsson, Gunlög. 2010. Object shift and optionality. An intricate interplay between syntax, prosody and information structure. *Working Papers in Scandinavian Syntax* 86.
- Larsson, Ida and Björn Lundquist. 2014. 'Objektsplacering vid partikelverb i norska dialekter och äldre svenska' in Janne Bondi Johannessen & Kristin Hagen (eds.) *Språk i Norge og nabolanda. Ny forskning om talespråk*. Oslo: Novus.
- Svenonius, Peter. 2002. 'Subject positions and the placement of adverbials.' In Peter Svenonius (ed.) *Subjects, Expletives and the EPP*: 201-242. New York: Oxford University Press.
- Thráinsson, Höskuldur. 2001. 'Object Shift and scrambling,' in Mark Baltin and Chris Collins (eds.) *The Handbook of Contemporary Syntactic Theory*: 148-202. Oxford: Blackwell.
- Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. New York: Oxford University Press.
- Holmberg, Anders and Christer Platzack. 1995. *The Role of Inflection in Scandinavian Syntax*. New York & Oxford: Oxford University Press.
- Westergaard, Marit. 2011. Subject positions and information structure: The effect of frequency on acquisition and change. *Studia Linguistica* 65 (3), 299–332
- Westergaard, Marit, Øystein Vangsnes & Terje Lohndal. 2017. Variation and change in Norwegian wh-questions: The role of the complementizer som. *Linguistic Variation* 17(1): 8-43.
- Wiklund, Anna-Lena, Gunnar H. Hrafnbjargarsson, Thorbjörg Hróarsdóttir, Kristine Bentzen. 2007. Rethinking Scandinavian verb movement. *Journal of Comparative Germanic Linguistics* 10: 203–233.