

Databases of elicited speech for analysis of syntactic and prosodic variation within and across speakers and dialects.

Björn Lundquist

presenting work done in collaboration with Ida Larsson, Paulina Lyskawa, Jade Sandstedt, Maud Westendorp, Eirik Tengesdal, Bror Magnus Sviland-Strand, Anders Nøklestad and Nathan Young

June 30th 2023

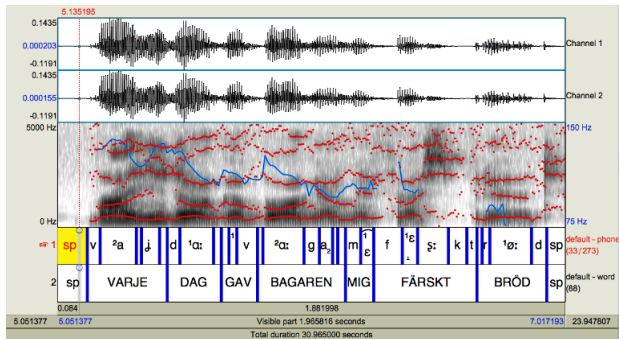
The Nordic Word order Database

- Database (available online <https://tekstlab.uio.no/nwd/>) containing elicited sentences from all the North Germanic languages, covering 10 core syntactic variables.

- (1)
 - a. **Subject shift:** Subject (Pro/DP) – Negation/Adverb
 - b. **Long object shift:** Subject (DP) – light pronominal object
 - c. **Reflexive Long object shift:** Subject (DP) – Refl. object
 - d. **“Long” particle shift:** Subject (Pro/DP) – Verb Part.
 - e. **Object shift:** Object (Pro/DP) – Neg/Adv
 - f. **Particle Shift:** Object (Pro/DP) – Verb Part.
 - g. **Embedded V2:** Fin. verb – Adv. in non-factive emb.clause
 - h. **V-to-I movement:** Fin. verb – Adv. in embedded quest.
 - i. **V3 in question:** Fin. verb – Subject (Pro/DP). in questions.
 - j. **V3 after preverbal adverbs:** Fin. verb – Adverb in Main clauses@
 - k. (+ comp, in embed, questions, articles, gender, prosody)

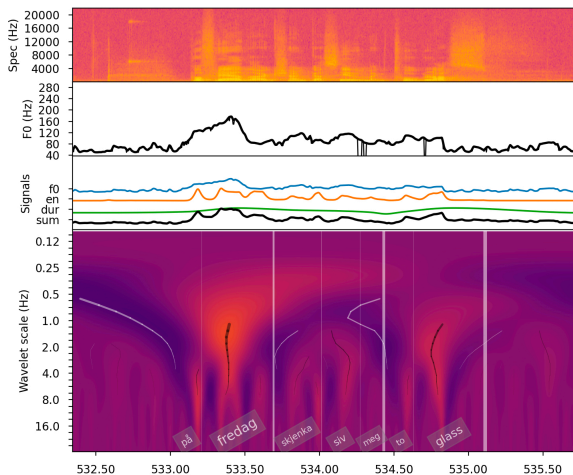
- Containing appr. 90.000 sentences, tagged for word order, with linked sound files.
- Goal: at least 10 items per variable and speaker, and at least 20 speakers per dialect, to establish if variation is within/between speakers/dialect/ languages. (<https://journals.uio.no/NALS/article/view/7529>)
- Various elicitation methods (spoken or written)

The acoustic analysis



- Forced alignment of the elicited material (Nathan Young: NoFA)
- Temporal measures: Onset latencies, total speech duration (segment, word and sentence level).
- Acoustic measures: f_0 , amplitude, length (word or segment) and formants.

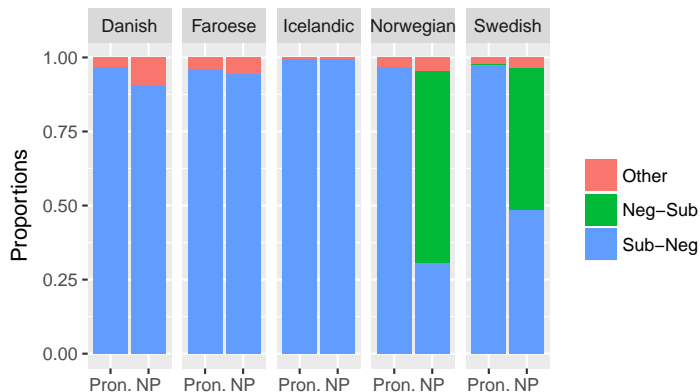
The acoustic analysis: Wavelet prosody toolkit, Suni et al. 2017



Boundary strength and word prominence using Wavelet Prosody Toolkit (e.g., boundary strength measures for each word). Hopefully we can use these value to (semi-)automatically obtain the prosodic structure.

Variation within and between speakers

- (2) Studenten åkte inte hem igår.
student.def went not hem yesterday
"The student will not go home tomorrow"
- (3) Igår... åkte {inte} studenten {inte} hem
yesterday went (not) student.def (not) home



Variation and underspecification – two schools

- 1 Probabilistic grammar: everything is underspecified, no "rules" in language.
- 2 The generative tradition: Strict mapping from form to meaning/structure. Two different word orders have different structures (and usually different meanings). If no meaning/structure differences can be detected, the two word orders belong to different grammars.

Mappings from an underlying syntactic structure to a linear order may be partly underspecified. RQs: (1) the architecture of grammar – how to capture both rigid and underspecified mappings from underlying structure to linear order. (2) Why do we get the categorical patterns at all?

Underspecified mappings from syntax to prosody – Particle stress

- a. Jón (gikk út) i hágen – Particle/phrasal accent
- b. Jón (gìkk ut) i hágen – Compound accent
- c. Jón (gíkk út) i hágen – Double accent
"Jon went out in the garden"

Tengesdal, Lundquist and Larsson (in proc): All three realizations are found within speakers, as opposed to e.g. true compounds (compound accent) and "phrasal verbs" (*kjøpe hus* – "buy house", phrasal accent).

Strict restriction on compound realization

- a. Jón (gikk út) i hágen – Particle/phrasal accent
- b. Jón (gìkk ut) i hágen – Compound accentt
- c. Jón (gíkk út) i hágen – Double accent

- Structural verb-particle relationship (no prepositional particle, no modal+null verb):

- (4) *Jón (gìkk ut) døra. (only phrasal or double acc.)
"Jon walked out the door."
- (5) *Jón (må ut) i hágen. (only phrasal or double acc.)
"Jon must (go) out in the garden"
- (6) *Da gikk mannen som (skrèv ut) i hagen.
Then, the man who was typing went out into the garden.

Strict restriction on compound realization

- a. Jón (gíkk út) i hágen – Particle/phrasal accent
- b. Jón (gìkk ut) i hágen – Compound accent
- c. Jón (gíkk út) i hágen – Double accent

- Phonological interveners: if an intervening element has to realize its lexical tone/accent, compound accent is impossible:

(7) *Da (gìkk (Jón) ut) i hágen – only phrasal or double.

(8) *Jón (gìkk (òfte) ut) i hágen – only phrasal or double.

Factors that probabilistically affect compound realization

- a. Jón (gikk út) i hágen – Particle/phrasal accent
 - b. Jón (gìkk ut) i hágen – Compound accent
 - c. Jón (gíkk út) i hágen – Double accent
- Number of intervening non-accented elements. (Sentence adverbials/particles pronouns.) Compare compounds, where length does not have this effect.

(9) Da (gìkk han ut) i hágen
then went he out in garden

(10) Da (kàsta jeg ham ut)
then threw I him out

(11) Jón (gìkk jo ut) i hágen.
John went M.PART out in garden

(12) Da (kàsta jeg ham ju ut)
then threw i him M.PART out

Factors that probabilistically affect compound realization

- a. Jón (gikk út) i hágen – Particle/phrasal accent
 - b. Jón (gìkk ut) i hágen – Compound accent
 - c. Jón (gíkk út) i hágen – Double accent
- The semantic relationship between verb and particle/collocation strength between verb and particle: compound accent more likely if verb and particle are a lexicalized unit, (13-a) compared to (13-b) (again compare with compounds)

(13) Da (gà han op)
then have he up
"Then he gave up"

(14) Da (spràng han ut)
then ran he out
"Then he ran out"

Factors that don't affect the likelihood of compound realization

- a. Jón (gikk út) i hágen – Particle/phrasal accent
 - b. Jón (gìkk ut) i hágen – Compound accent
 - c. Jón (gíkk út) i hágen – Double accent
- (Final) syntactic structure: It doesn't matter if the verb stays in the verb phrase (embedded clause, simple tense or complex tenses, (15-a)) or surfaces in V2 position (and then not obviously making up a syntactic constituent with the particle, (15-b)):

(15) Jon skal [_{VP}(gå út) i hágen]
Jon will go out in garden
"Jon will go out into the garden"

(16) Jón [_{CP}(gíkk [_{VP} ut) i hágen]]
Jon went out in the garden
"John went out into the garden."

Factors that don't affect the likelihood of compound realization

- a. Jón (gikk út) i hágen – Particle/phrasal accent
 - b. Jón (gìkk ut) i hágen – Compound accent
 - c. Jón (gíkk út) i hágen – Double accent
- Syntactic function of an intervening pronoun: an intervening subject (17-a) and object (17-b) reduces the likelihood of compound accent to the same extent:

(17) Jón har [_{VP}(kàsta han ut) i hágen]
Jon has thrown him out in garden.DEF
"Jon has thrown him out in the garden"

(18) Da [_{CP}(kàsta _{TP}[han [_{VP}ut) hunden]]]
Then threw he out dog.DEF
"Then he threw out the dog."

Underspecified mapping from syntax to Prosody

- (19) Jón har [_{VP}(kàsta han ut) i hágen]
Jon has thrown him out in garden.DEF
"Jon has thrown him out in the garden"
- (20) Da [_{CP}(kàsta _{TP}[han [_{VP}ut) hunden]]]
Then threw he out dog.DEF
"Then he threw out the dog."

Prosody/PF cares about the linear order and the "deep" structural relation between verb and particle, but not the (final) syntactic configuration.

- Strict: Underlying verb-particle structural relation ("syntactic function") and lexical phonological representation.
- Probabilistic: all the typical "usage" factors (length of unaccented interveners, verb-particle collocation...) – effects of production planning.

(but remember, for other phenomena/dialects, mappings might be strict)

Experimental approaches to Syntactic Optionality

- Main Q: Why do strict categorical rule systems evolve in languages?
- How: study processing and acquisition of variable and categorical rules by both adults and children, in three closely related languages (Norwegian, Swedish and Danish) which show varying degrees of optionality.
- Hypothesis to be tested:
 - 1 categoricity arises from an innate or early bias against synonymy/optionality (e.g., Clark 1987);
 - 2 categoricity evolves gradually over generations due to children's tendency to over-regularise variation (possibly due to memory limitations, Newport 1985, Yang 2012);
 - 3 templatic patterns arise due to processing demands in adult language production (MacDonald 2013, Levinson 2017).

By studying processing (production/comprehension) and acquisition of variable and non-variable grammars in the context of closely related languages, we can identify where preferences for regular systems arise.

- 1 are there processing benefits (or costs) with rigid mappings from meaning to form (here, word order)?
- 2 is the L1 language learner pre-disposed to categorical meaning-to-form mappings, or do categorical tendencies develop over time (in individuals)?
- 3 How is word order optionality modelled in a theoretical framework? What is the locus of optionality, and how do intragrammatical factors (prosody, semantics) influence variable word order patterns?

Underspecified (and partly arbitrary) mapping from syntax to constituent order

Prime **Läraren** kommer **inte** att hjälpa **mig** med läxan.
teacher.def will not to help me with homework.DEF

Target Igår hjälpte...
yesterday helped...

Swedish, Norwegian and Danish

Linear order	Verb	order	Swe.	Norw.	Dan.
S>N, O>N, S>O	helped	teacher me not	OK	OK	OK
N>S, N>O, S>O	helped	not teacher me	OK	OK	*
S>N, N>O, S>O	helped	teacher not me	OK	*	*
S>N, O>N, O>S	helped	me teacher not	OK	*	*
N>S, O>N, O>S	helped	me not teacher	OK	*	*
N>S, O>N, O>S	helped	not me teacher	OK	*	*

"Dumb" word order restrictions: Object shift, Subject shift, Long Object shift

- (21) a. Læreren hjalp *meg* ikke med leksene.
b. ??Læreren hjalp ikke *meg* med leksene.
The teacher helped not me with the homework
- (22) a. I går hjalp ikke *læreren* studenten sin.
b. I går hjalp *læreren* ikke studenten sin.
Yesterday helped the teacher not his student.
- (23) a. *I går hjalp *meg* ikke *læreren*.
b. I går hjalp *læreren meg* ikke.
c. I går hjalp ikke *læreren meg*.

"Dumb" Linearization statements in Norwegian:

- (24) a. Subject > Object (No Long OS)
b. ProObj > Neg (Object Shift (OS))
c. (Neg > NPSub or NPSub > Neg) (Var. NP SubShift)

Subject shift and prosodic reusability

The effect of elicitation method:

- (25) Studenten (åkte inte) hem igår.
student.def came nor too late yesterday
- (26) Studenten kommer inte att åka hem
student.def will not INF go home

Elicitation method in (25) triggers more Neg-Sub than elicitation method in (26) (see Larson & Lundquist 2022, Lundquist & Tengesdal 2022).
If speakers can re-use the prosodic chunks, they will do that.

Prosodic constituents affects choice of word order

In two of our cases, the prosodic constituency will differ depending on choice of word order:

1. SubjNP-Neg: (a) (Igår) (**kom inte**) (*läraren*) (hem) – (b) (Igår) (*kom läraren*) (*inte*) (hem)
2. ObjNP-Part: (a) (*Vaktene*) (**kastet ut**) (*studenten*) – (b) (*Vaktene*) (*kastet*) (*studenten*) (*ut*)

Word order choice of type (a) is more frequent the tighter the prosodic unity is between the verb and the negation/particle in the self-produced background/prime sentence, as estimated by prominence and boundary values from WPT (glmers, with word order as dep. var, and prominence/boundary as predictors):

Subj-Neg: $\beta=0.86$, *St.err*=0.37, $p=0.02$; Obj.Part: $\beta=0.37$, *St.err*=0.13, $p=0.014$

That is, the relative order between two words in the prime sentence is more likely to be maintained in the the target if the two words form a tight prosodic unity (“a prosodic word”) in the prime.

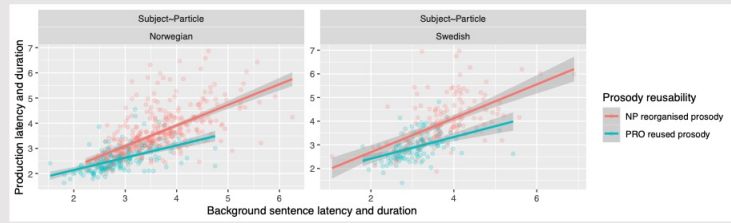
Prosodic reorganization slows you down

There is only one condition where prosodic reorganization is forced ,i.e., where prosodic words cannot be re-used, and this is the subject-particle condition:

(Läraren) (gav upp) (igår) – (Igår) (gav läraren) (upp)

(Teacher.def) (gave up) (y.day) --(Y.day) (gave teacher.def) (up)

This is the only condition where we find a clear increase in speech times (compare to re-usable prosodic structure).



Object shift: Strict structure-mappings, later over-ruled by PF?

(27) Hon hjälpte mig inte.
she helped me not
'She did not help me.'

(28) *Hon hjälpte inte mig.
she helped not me
'She did not help me.' (Swedish)

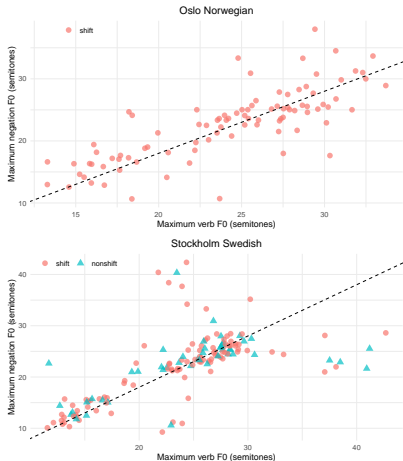
OS: prosodic incorporation of prosodically weak pronominal objects, which preferably incorporate into preceding non-adverbial prosodic hosts. This process can remain flexible in certain Scandinavian dialects that have tone accent distinctions since their prosody provides more variants of prosodic incorporation which non-tonal varieties lack.

- (29) Predictions of Erteschik-Shir, Josefsson, and Köhnlein proposals
- Shifted word order should be unacceptable with non-incorporating focally accented verbs.
 - Shifted word order should be obligatory with non-incorporating focally accented negation.

Finding from prosodic analysis: All possible combinations of shifted and unshifted word order and incorporation and non-incorporation of weak objects are licit, showing that there is no necessary link between prosodic incorporation and OS. (Lyskawa et al. 2022: *Successes and shortcomings of phonological accounts of Scandinavian Object Shift*)

Downstep independent of OS

Capitalizing on the proposal by gussenhoven2004 phonology that downstep intonation is driven by a L(ow)-tone element intervening between two Hs, Hosono proposes that light pronouns in Scandinavian can provide such L-elements. Thus, when placed in the right configuration ($H_{\text{VERB}}L_{\text{PRONOUN}}H_{\text{NEGATION}}$), Hosono claims that the pronoun triggers downstep between the adjacent Hs.



In sum, we do not find corroborating results that shift entails downstep in sentences where shift occurs. (...) If there are pathways for downstep other than shift, shift is not necessary for a information-structural neutral interpretation.

Final slide

- Good idea to carry out large scale elicitation studies.
- Challenges in analysing the acoustic/prosodic data
- Challenges in making the phonetic data available. Currently, our database can only be searched for syntactic data + metadata.