

George Starostin (RSUH / RANEPa, Moscow)



«Micro»-lexicostatistics: the Scandinavian case

[The 11th Annual Sergei Starostin Memorial Conference on Comparative-Historical Linguistics](#)

Russian State University for the Humanities, March 24-25, 2016

Preliminary notes

- **In general:** Germanic data, and Scandinavian data in particular, very important for the purposes of glottochronological calibration
- **Bergsland & Vogt 1962:** Use Scandinavian data to conclusively disprove the validity of Swadesh's glottochronological constant (14% replacements on the 100-item wordlist per millennium)
- **Sergei Starostin 1989:** Reanalysis of the data shows that replacement rates across Scandinavian languages are comparable if borrowings ("externally driven change") are excluded from calculations
- **Problem 1:** Most existing wordlists for Germanic languages suffer from various degrees of imprecision. Accuracy and consistency rarely achieved to the degree that is necessary for such closely related languages
- **Problem 2:** No attempts to estimate the number of lexical replacements from Proto-Germanic to various subsequent stages (e. g. Proto-Scandinavian, Proto-West Scandinavian, Proto-East Scandinavian, etc.)

What we did:

- **Compilation:** New wordlists for five modern Scandinavian languages (Swedish, Bokmål Norwegian, Danish, Icelandic, Faroese), as well as for 13th century Old Norse and one outlier (4th century Gothic). Based on dictionaries and textual corpora
- **Data processing:** Items were selected, transcribed, and annotated (including discussion of synonymy wherever possible) in accordance with the common standards of the Global Lexicostatistical Database project (<http://starling.rinet.ru/new100>)
- **Calculations:** Glottochronological matrix and genealogical tree produced according to Sergei Starostin's revised formula (Swadesh formula unusable in this case)
- **Etymological support:** Approximate number of replacements manually calculated from Proto-Germanic to Proto-Scandinavian and its further branchings

Example of a text-converted non-problematic item:

21. EAR

Gothic *aus-o:* (1), Old Norse *eir-a* (1), Icelandic *ei:r-a* {*eyra*} (1), Faroese '*oi:r-a* {*oyra*} (1), Bokmal Norwegian *ø:r-ε* {*øre*} (1), Danish '*ö:ɹ-ə* {*øre*} (1), Swedish '*ö:r-a* {*öra*} (1).

References and notes:

Gothic: Balg 1887: 37. Neuter gender; *n*-stem. Cf. Mk. 4:9: *saei habai ausona hausjandona, gahausjai* "he that has ears to hear, let him hear" [Ulfilas 1896: 25].

Old Norse: Cleasby & Vigfusson 1874: 135; Zoega 1910: 120; De Vries 1962: 107. Neuter gender. Plural form: *eir-u*.

Icelandic: Haraldsson 1996: 769; Berkov 1962: 151. Neuter gender. Cf. *hlust* 'ear (applied to certain animals, e. g. bears, birds, etc.) / ear canal' [Berkov 1962: 292].

Faroese: Young & Clewer 1985: 432. Neuter gender.

Bokmal Norwegian: Arakin 2000: II, 476; Berkov 2006: 1042. Definite form: *øre-t*.

Danish: Krymova et al. 2000: 802; Harrit & Harrit 2002: 462. Definite form: *øre-t*.

Swedish: Marklund-Sharapova 2007a: 918; Marklund-Sharapova 2007b: 748. Definite form: *ör-at*. Plural: *ör-on*.

Example of a text-converted problematic item:

4. BELLY

Gothic *wamb-a* (1), Old Norse *kvið-r* (2), Icelandic *kvið-yr* {*kviður*} (2), Faroese *buk-yr* {*búkur*} (3), Bokmal Norwegian *måg-ε* {*mage*} (4), Danish *m'avə* {*mave*} (4), Swedish *m'v:gə* {*mage*} (4).

References and notes:

Gothic: Balg 1887: 528. Fem. gender. Most of the contexts refer to the semantics of 'womb' (referring primarily to the Mother of God), but there are a few passages that confirm the general meaning 'belly' as well, cf. Mk. 7:19: *ni galeiþiþ imma in hairto, ak in wamba* "it does not enter in his heart, but in his belly" [Ulfilas 1896: 32], etc. It is somewhat hard to establish the difference between *wamb-a* and its quasi-synonym *qīþus*, glossed in [Balg 1887: 229] as 'womb; stomach'. The latter, however, never translates Greek *κοιλια* 'belly', and is at least once encountered in the precise meaning 'stomach' (body organ rather than body part), cf. I Tim. 5:23: *weinis leitil brukjais in qīþaus þeinis* "take a little wine for your stomach" [Ulfilas 1896: 204]. It may, therefore, be surmised that *wamba* referred primarily to the inside part of the body (which is the required Swadesh meaning', whereas *qīþus* had the polysemous semantics of 'womb / stomach' ('organ inside the belly').

Old Norse: Cleasby & Vigfusson 1874: 364; Zoega 1910: 254; De Vries 1962: 338. Masculine gender. Polysemy: 'belly / womb'. Distinct from *magi* [Cleasby & Vigfusson 1874: 408], which is predominantly 'stomach' or 'maw', and from *vomb* [Cleasby & Vigfusson 1874: 722] which seems to be a "vulgar" equivalent of 'belly' (Cleasby: "mostly in a low sense, especially of beasts"). Overall, there is some significant contextual overlap between all the three words, but the underlying opposition of 'belly ~ womb', 'stomach', and 'belly (vulg.)' seems to suggest *kvið-r* as the most eligible candidate (*contra* [Bergsland & Vogt 1962: 117], where 'belly' is still rendered as *magi*, whereas *kvið-r* is glossed as 'less inclusive; womb' - this judgement is not supported well by the data in Cleasby's dictionary, but, perhaps, a more detailed scrutiny is required).

Icelandic: Haraldsson 1996: 191; Berkov 1962: 382. Distinct from *magi* 'stomach' [Berkov 1962: 430] (also used in contexts of 'stomach ache'); from *vömb* 'belly' (vulgar equivalent) [Berkov 1962: 875].

Faroese: Young & Clewer 1985: 67. Meaning glossed as 'abdomen, stomach, belly; waist; (ballad) body'. The partial synonym *magi* [Young & Clewer 1985: 372], glossed as 'stomach, belly', seems to be primarily referring to 'stomach' as an organ or to the interior part of the belly. Cf. also *kviður* 'abdomen, stomach' [Young & Clewer 1985: 327], although this word is never given as the default Faroese equivalent for 'belly' in any of the sources. On the whole, this seems like one of the very few Swadesh items to reflect a lexicostatistical difference between Icelandic and Faroese.

Bokmal Norwegian: Arakin 2000: I, 520; Berkov 2006: 229. "Traditional Bokmål": *mave* (Danish-influenced form). Polysemy: 'belly / stomach'. Distinct from *buk* 'belly (of animal)', used also of persons in a vulgar sense [Arakin 2000: I, 146; Berkov 2006: 229].

Danish: Krymova et al. 2000: 418; Harrit & Harrit 2002: 105. Polysemy: 'stomach / belly / bowels'. Another synonym is *bug* [Krymova et al. 2000: 103], usually seen as a more vulgar (and statistically less frequent) equivalent (see the same situation carried over to Bokmål Norwegian). The word *abdomen*, listed as another equivalent for 'belly' in [Harrit & Harrit 2002: 105], is a specialized anatomical term.

Swedish: Marklund-Sharapova 2007a: 436; Marklund-Sharapova 2007b: 154. Polysemy: 'stomach / belly'. Definite form: *mag-e-n*. Plural: *mag-ar*. Another synonym is *buk* [Marklund-Sharapova 2007a: 95], but it seems to be confined to the "vulgar" register, as in Danish.

Investigating lexicostatistical replacements: From Proto-Germanic to Proto-Scandinavian

I. Proto-Germanic retentions in Proto-Scandinavian

I.1. Items that remain unchanged in every Scandinavian language as well as in Gothic and are unquestionably reconstructible for Proto-Germanic:

Out of the first 50 items («more stable part»): *ashes, bird, black, blood, die, dog, drink, ear, eye, foot, hand, head, hear, heart, horn, I, mouth, name, new, night, one, rain, star, stone, sun, thou, tongue, tooth, two, water, we, what, who* (33)

Out of the second 50 items («less stable part»): *all, bite, breast, burn, cold, come, earth, fish, full, give, good, knee, know, lie, long, many, neck, person, red, road, root, see, seed, sit, small, stand, white, woman* (28)

Investigating lexicostatistical replacements: From Proto-Germanic to Proto-Scandinavian

I. Proto-Germanic retentions in Proto-Scandinavian

I.2. Items that remain unchanged in every Scandinavian language and are not attested in Gothic, however, are well attested in West Germanic and are also unquestionably reconstructible for Proto-Germanic:

claw(nail), dry, egg, fat, feather, fly, green, liver, louse, smoke, yellow

Adding them to items in I.1 yields at least **72/100** unambiguous retentions in Proto-Scandinavian from Proto-Germanic.

Investigating lexicostatistical replacements: From Proto-Germanic to Proto-Scandinavian

1.3. Items that are replaced in at least some Scandinavian languages or in Gothic. However, strong arguments from Gothic, West Germanic, and/or external IE data allow to unambiguously posit them as Proto-Scandinavian items:

<i>bone</i>	(+ West G.; Gothic unknown)
<i>eat</i>	(+ West G.; + IE; Gothic vulgarism)
<i>go</i>	(+ West G.; Gothic shift to more figurative meaning)
<i>man</i>	(+ Gothic; + West G.)
<i>moon</i>	(+ Gothic; + West G.; + IE)
<i>nose</i>	(+ West G.; + IE; Gothic unknown)
<i>sand</i>	(+ West G.; + IE?; Gothic replacement)
<i>say</i>	(+ West G.; + IE?; Gothic replacement)
<i>sleep</i>	(+ IE; Gothic and West G. shared replacement)
<i>swim</i>	(+ West G.; Gothic unknown)
<i>that</i>	(+ Gothic; + West G.; + IE?)
<i>this</i>	(+ Gothic; + West G.; + IE?)
<i>tree</i>	(+ IE; Gothic and West G. shared replacement)
<i>warm</i>	(+ West G.; Gothic unknown)

Investigating lexicostatistical replacements: From Proto-Germanic to Proto-Scandinavian

I.4. Probable, but somewhat questionable retentions in Proto-Scandinavian:

cloud (ON *ský*, etc.; cf. OE *scēo* 'cloud';
altern.: Gothic *milx-ma* = Swedish *mol-n*?)

hair (ON *hár*, etc.; + West G.; against Gothic *tagl*)

mountain (Swedish *berg*, etc.; + West G.; + IE?;
against Gothic *fairguni*)

skin (Icel. *húð*, etc.; + West G.; + IE?)

Adding I.3 and I.4 (18 additional items) to items in I.1 and I.2 yields at least **90/100** retentions in Proto-Scandinavian from Proto-Germanic (4 of them slightly ambiguous).

Investigating lexicostatistical replacements: From Proto-Germanic to Proto-Scandinavian

I.5. Unquestionable replacements in Proto-Scandinavian (confirmed by IE data):

<i>fire</i>	(*fō-n ~ *fū-r → *eld-r /‘flame ~ spark’ → ‘fire’)
<i>meat</i>	(*mimza- → *kjot / ‘intestines?’ → ‘meat’)
<i>not</i>	(*ne → *ei-gi /shift from an emphatic time adverb/)

I.6. Possible replacements in Proto-Scandinavian

<i>bark</i>	(derived from ‘birch’; no clear Proto-Germanic equivalent)
<i>big</i>	(original *mikil- still found in ON /archaism?/; main word is *stōr-)
<i>belly</i>	(no single clear Proto-Germanic or Proto-Scandinavian term)
<i>kill</i>	(no clear Proto-Germanic; Scandinavian *drepa ← ‘to strike’)
<i>leaf</i>	(ON <i>blað</i> , etc., originally ← ‘blade’; cf. ON <i>lauf-blað</i> ‘a single leaf’, with old *lauba- generalized as the collective ‘foliage’)
<i>round</i>	(no single clear Proto-Germanic or Proto-Scandinavian term; possibly PS *kring-l- ‘round’, ‘ring’)
<i>tail</i>	(no single clear Proto-Germanic term; ON <i>hali</i> , etc., finds no Germanic or IE support in the meaning ‘tail’)

Investigating lexicostatistical replacements: From Proto-Germanic to Proto-Scandinavian

Altogether:

Approximately 90 retentions vs. 10 innovations (10% replacements).

Cf. the structure of the Gothic wordlist:

- 71 retentions from Proto-Germanic
- ≈10 innovations
- 19 lacunae
- ≈12% replacements

Investigating lexicostatistical replacements: From Proto-Scandinavian to Modern West Scandinavian

II.1. Innovations in Proto-West Scandinavian (ancestor to Old Norse, Icelandic, Faroese)

[1] **man-* 'man (male human being)' → **karl*

[2] **gǫng-* 'go' → **far-a* (**gǫng-* still preserved in more figurative meanings)

II.2.1. Innovations in 13th century Old Norse (ancestor to Icelandic)

[1] **māni* 'moon' → ON *tungl* (cf. Faroese *máni*; in ON, the original word is an archaism)

II.2.2. Innovations in Modern Icelandic (from Old Norse)

[1] *et-a* 'eat' → Icelandic *borða*

[2] *svim-a* 'swim' → Icelandic *synda*

[3] *nas-ar* 'nose' → Icelandic *nef* '*beak'

(terms already synonymous in ON, but not in Proto-West Scandinavian, cf. Faroese *nøs* 'nose')

II.2.3. Innovations in Faroese (from Proto-West Scandinavian)

[1] *búkur* 'belly' (old vulgarism)

[2] *rund-ur* 'round' (borrowing)

[3] (?) *he-sin* 'this' (← **ðe-sin* by analogy with *ha-sin* 'that?')

Investigating lexicostatistical replacements: From Proto-Scandinavian to Modern East Scandinavian

III.1. Innovations in Proto-East Scandinavian (ancestor to Danish, Swedish)

- [1] *? 'belly' → **mage* '*stomach'
- [2] **kringl-* 'round' (?) → **rund-* (Germanic borrowing)
- [3] **hi-n-* 'that' → **de-n-*

III.2.1. Innovations in Danish

- [1] *ben* 'bone' > *knogle* (German borrowing)
- [2] *et-* 'eat' > *spise* (German borrowing)

III.2.2. Innovations in Swedish

- [1] *sky* 'cloud' → *moln*
- [2] *drep-* 'kill' → *döda*
- [3] *ikke* 'not' → *inte* (from **inget*) 'nothing'
- [4] *hali* 'tail' → *svans* (German borrowing)

Investigating lexicostatistical replacements: The case of Bokmål

IV. Lexicostatistical differences between Danish and Bokmål

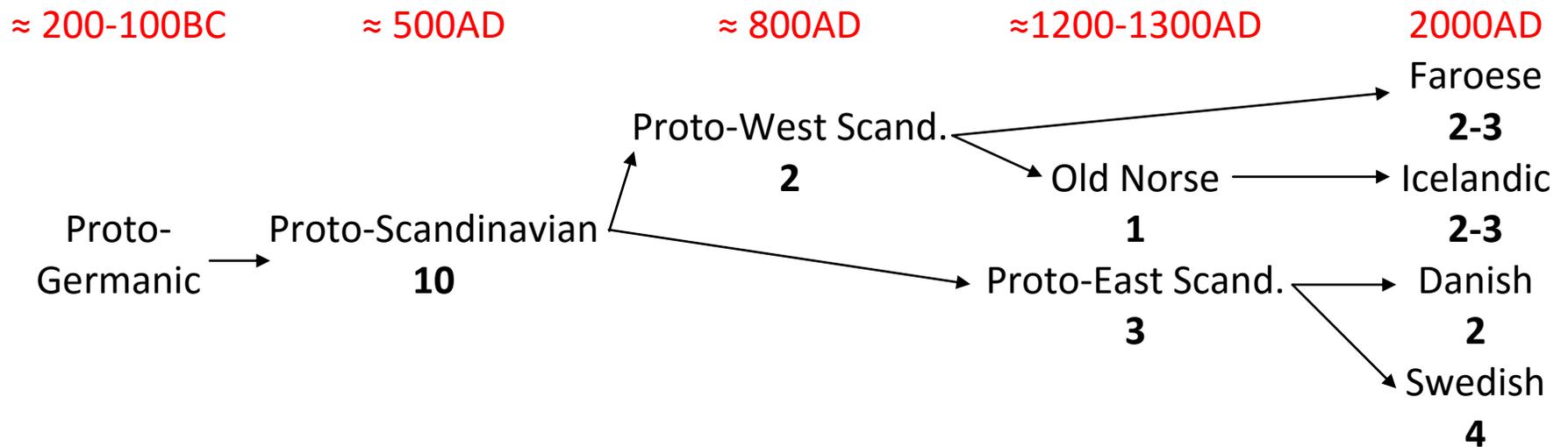
Word	Danish	Bokmål
'bone'	<i>knogle</i>	<i>bein ~ ben</i>
'mountain'	<i>bjerg</i>	<i> fjell</i> (West Sc. term) East Sc. <i>berg</i> also exists, but has more of a collective meaning

V. Some phonetic discrepancies between Danish and Bokmål

Word	Danish	Bokmål	"Traditional Bokmål"
'belly'	<i>mave</i>	<i>mage</i>	<i>mave</i>
'black'	<i>sort</i>	<i>svart</i>	<i>sort</i>
'smoke'	<i>røg</i>	<i>røyk</i>	<i>røk</i>
'stone'	<i>sten</i>	<i>stein</i>	<i>sten</i>

Conclusion: At least strictly lexicostatistically, Bokmål *is* a form of Danish. It makes more sense to count 'bone' and 'mountain' as borrowings from West Scandinavian dialects, and the other examples as secondary "Norwegisations" of Danish words, than to try and exclude "Danish borrowings" from lexicostatistical calculations.

Chronological summary of Scandinavian lexical innovations



Notes:

(1) German borrowings in Danish and Swedish are included in these figures.

(2) Datings reflect the presumed split of the taxon (e. g., «Proto-Germanic» is assumed to have split into Northern / East / West circa 200-100 BC, etc.).

Preliminary conclusions

(1) Lexical change between Proto-Germanic and Proto-Scandinavian occurred at about twice the rate of lexical change between Proto-Scandinavian and *all* its descendants. This is in accordance with Sergei Starostin's observations (1989) on gradual deceleration of change in the past 1000 years.

(2) No specific «Icelandic conservatism» to speak of: *all* Scandinavian languages exhibit comparable degrees of lexical conservatism (only Swedish is a little bit more innovative).

(3) On the whole, Proto-Germanic on the way to its modern Scandinavian descendants, over a little more than 2000 years of change, shows **15-17** lexical replacements.

Significantly different from comparable results for the time period between Classic Old Greek (4th century BC) and Modern Greek (\approx **38** replacements); or for the time period between Classical Chinese (5th-4th century BC) and Modern Chinese (\approx **32** replacements).

Roughly speaking, the «Swadesh constant» (0.14) makes more sense for Greek and Chinese, whereas the «Starostin constant» (0.05) makes more sense for Scandinavian.

What next?

(1) Further improvements of the dataset:

- clear up some ambiguities in already compiled lists
- add new lists for spoken West/East dialects (Norwegian, Danish, etc.)
- complete the Germanic database with West Germanic lists

(2) Add more datasets for languages with attested histories for comparison, to answer questions such as:

- is the «Scandinavian case» or the «Greek/Chinese case» the norm or the exception?
- what sort of sociolinguistic reality underlies the discrepancies?
- can the difference be effectively integrated into current glottochronological models?