

Florian Dolberg

Why ‘big data’ might not always be the best data in historical linguistics: a case study on the role of contact in structural changes of mediaeval English

It is widely-known that English underwent many drastic structural changes from Old English (OE) to late Middle English. Also widely-known is that Viking settlement-names at that time were very numerous, but confined to the Danelaw-area in northern and eastern England (cf. e.g. Stenton 1942; Cameron 1969: 75-86, 1971; Richards 1991: 33-36; Loyn 1994: 82-89; Crawford 2003: 59-60). And while it is generally uncontested that English borrowed quite a few and very basic words from Old Norse (ON) (cf. e.g. Hofmann 1955; Peters 1981; Thomason & Kaufman 1988; Townend 2002; Pons-Sanz 2007, 2013; Lutz 2012), there is much discord in the research-community regarding the role of ON/OE-contact in the structural changes mediaeval English underwent: outspoken affirmers of contact-induced or -driven grammatical change (e.g. McWhorter 2002; Townend 2002) face vehement deniers (e.g. Thim 2008, 2012; Lutz 2012).

This paper contributes to overcoming this stand-off by offering a case-study on English gender-development, a particularly puzzling instance of language-change: although more than a dozen studies on English gender-change precipitated since Körner’s (1888) initial treatise, this large body of research has produced “remarkably contradictory results” (Stenroos 2008: 451) regarding pathways of and reasons for gender-simplification. While a number of contributions found only chaos, i.e. absence of patterns (e.g. Lindelöf 1893, Ausbüttel 1904, Markus 1988), there seems to be none yet that explicitly views the phenomenon in the context of OE/ON contact.

These contradictory and/or negative results have two likely causes: one is methodological. Large-scale studies typically analyse miscellanies of documents of strikingly dissimilar regional provenance, text-type/genre, topic, and even age (cf. Stenroos 2008: 451), all of which can exert a considerable confounding influence (cf. e.g. Szmrecsanyi 2016), “making it difficult to trace coherent patterns” (Stenroos 2008: 451). Conversely, though to a similar effect, smaller studies that base on a single document have no way of disentangling particularities of the document at hand from general trends and characteristics. The other reason is too narrow a theoretical focus, i.e. investigating only one type of potential predictor – say, phonology – while ignoring all or most others – such as morphological, semantic, cognitive, and contact effects, thus obtaining a partial and distorted picture (cf. e.g. Enger 2013 for an eloquent critique of such approaches).

As historical language-data is inherently limited, and as “[h]istorical documents survive by chance, not by design” (Labov 1994: 11), Labov described historical linguistics as “the art of making the best use out of bad data” (1994: 11). This paper hence suggests to neutralise or at least minimise the confounding influences sketched above by carefully selecting documents comparable in terms of text-type, style, and topic, drawing (where possible) on interdisciplinary evidence to ascertain age and provenance. Dividing the database thus

obtained into a north-eastern, OE/ON-contact-exposed cohort and a south-western, essentially OE/ON-contact-free cohort shows contact to quicken, possibly cause gender-simplification, but also show this process to underlie phonological, morphological, semantic and cognitive factors as well. This shows that the confounding effect of dissimilar documents in large datasets can be minimised by choosing a smaller, but more homogenous database instead.

References:

- Ausbüttel, Erich. 1904. *Das persönliche Geschlecht unpersönlicher Substantiva einschließlich Tiernamen im Mittelenglischen*. Tübingen: Niemeyer.
- Cameron, Kenneth. 1969. *English Place Names*. London: Methuen.
- Cameron, Kenneth. 1971. "Scandinavian settlement in the territory of the Five Boroughs: the place-name evidence Part III, the Grimston Hybrids". In: Peter Clemoes & Kathleen Hughes (eds.). *England before the Conquest*. Cambridge: Cambridge University Press.
- Crawford, Barbara E. 2003. "The Vikings". In: Wendy Davies (ed.). *From the Vikings to the Normans*. 14-71. Oxford: Oxford University Press.
- Enger, Hans-Olav. 2013. "Inflectional change, 'sound laws' and the autonomy of morphology. The case of Scandinavian case and gender reduction". *Diachronica* 30(1), 1-26.
- Hofmann, Dietrich. 1955. *Nordisch-englische Lehnbezeichnungen in der Wikingerzeit*. Kopenhagen: Munksgaard.
- Körner, Karl. 1888. *Beiträge zur Geschichte des Geschlechtswechsels der englischen Substantiva. Inaugural-Dissertation*. Greifswald: Julius Abei.
- Labov, William. 1994. *Principles of Linguistic Change, Volume I: Internal Factors*. Oxford: Blackwell.
- Lindelöf, Uno. 1893. "Beiträge zur Kenntnis des Altnordhumbrischen". In: Uno Lindelöf (ed.). *Mémoires de la Société Néo-Philologique a Helsingfors I*. Helsingfors: Imprimerie centrale de Helsingfors.
- Loyn, Henry R. 1994. *The Vikings in Britain*. Oxford: Blackwell.
- Lutz, Angelika. 2012. "Norse influence on English in the light of general contact linguistics". In: Irén Hegedüs & Alexandra Fodor (eds). *English Historical Linguistics 2010: Selected Papers from the sixteenth International Conference on English Historical Linguistics (ICEHL 16), Pécs, 23-27 August 2010*. Amsterdam/Philadelphia: John Benjamins.
- Markus, Manfred. 1988. "Reasons for the loss of gender in English". In: Dieter Kastovsky & Gero Bauer (eds.). *Luick revisited: Papers read at the Luick Symposium at Schloss Liechtenstein. 15-18.9.1985*. Tübingen: Narr.
- McWhorter, John H. 2002. "What happened to English?". *Diachronica* 19(2), 217–272.
- Peters, Hans. 1981. "Zum Skandinavischen Lehngut im Altenglischen". *Sprachwissenschaft* 6, 85-124.
- Pons-Sanz, Sara M. 2007. *Norse-derived vocabulary in late old English texts: Wulfstan's works, a case study (3rd ed.)* (= NOWELE Supplement Series, 22). Odense: University Press of Southern Denmark.
- Pons-Sanz, Sara M. 2013. *The lexical effects of Anglo-Scandinavian linguistic contact on old English*. (Studies in the Early Middle Ages). Brepols: Turnhout.
- Richards, Julian D. 1991. *Viking Age England*. London, New York, Sydney, Toronto: BCA.
- Szmrecsanyi, Benedikt. 2016. About text frequencies in historical linguistics: disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory* (special issue, ed. by Martin Hilpert & Hubert Cuyckens).

- Stenroos, Merja. 2008. "Order out of chaos? The English gender change in the Southwest Midlands as a process of semantically based reorganization". *English Language and Linguistics*, 12.3: 445-473.
- Stenton, Frank Merry. 1942. "The historical bearing of place-name studies: the Danish settlement of eastern England". *Transactions of the Royal Historical Society* 42: 1-24.
- Thim, Stefan. 2008. "The Rise of the Phrasal verb in English: A Case of Scandinavian Influence?" In: Klaus Stiersdorfer (ed.). *Anglistentag 2007, Münster: Proceedings*. Trier: Wissenschaftlicher Verlag.
- Thim, Stefan. 2012. *Phrasal Verbs: The English Particle Construction and Its History*. Berlin: Mouton de Gruyter.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Townend, Matthew. 2002. *Language and history of Viking age England: linguistic relations between speakers of Old Norse and Old English*. Turnhout: Brepols.