

Investigating interactional syntactic change in Middle English: Insights from visual analytics

Schätzle, Christin *University of Konstanz* christin.schaetzle@uni-konstanz.de

Booth, Hannah *University of Konstanz* hannah.booth@uni-konstanz.de

The idea that language change results from multiple interacting factors is long-standing in historical linguistics (e.g. Labov 1963, Malkiel 1967, Weinreich et al. 1968). Change can be the product of interacting language-internal (i.e. system-driven) and language-external (i.e. socio-political) factors, or result from multiple interacting, exclusively language-internal factors. *Syntactic* change is taken to interact with changes at other linguistic dimensions (e.g. phonology, morphology, semantics) and has thus been labelled an interface phenomenon (Keenan 1994, Longobardi 2001). Moreover, interactional change *within* the syntactic system has been addressed from various theoretical perspectives, e.g. via underlying parametric change in the Principles and Parameters approach (e.g. Kroch 1989, Lightfoot 2013), and in the usage-based paradigm under the notion of ‘multiple source constructions’ (van de Velde et al. 2013).

Alongside increased interest in interactional syntactic change, corpus-based methodologies specific to the problem have become increasingly sophisticated, yielding many novel findings (e.g. Hilpert and Gries 2016, Pintzuk et al. 2017). Nevertheless, the tools for investigating syntactic interactions in diachronic corpus-data remain insufficient. The standard procedure is to extract patterns from corpora via programming scripts or specific query tools, in order to generate data tables containing co-occurrence frequencies and statistical calculations. Finding patterns in such tables is challenging; various feature interactions have to be examined across several tables while taking into account a temporal component. Statistical testing can be useful for quantifying change and validating given hypotheses. However, it is less suited for uncovering syntactic change *per se*, since the precise factors involved cannot always be anticipated. Additionally, historical data is naturally sparse, which in turn might render statistical significances delusive.

In this paper, we show how visual analytics (Keim et al. 2008) offers a fruitful approach to investigating interactions between changing syntactic phenomena over time via the HistoBankVis visualisation system for historical linguistics (Schätzle et al. 2017, Schätzle et al. 2019). As a showcase, we examine word order change underway during the Middle English (ME) period (c.1100-1500), specifically the loss of the verb-second (V2) constraint and the emergence of S(ubject)-V(erb)-O(bject) word order. Various factors have been claimed relevant for this change (e.g. Los 2009, van Kemenade 2012), but the precise nature of their interactions remains elusive.

Early English texts exhibit a good deal of variation with respect to clausal word order. Old English (OE) is often characterised as a V2 language (e.g. Holmberg 2015), but there is evidence that V2 was not fully consolidated. A relevant factor is the category of

clause-initial constituent (van Kemenade 1987, Pintzuk 1999). In clauses with an initial *wh*-phrase, negator or discourse adverb (e.g. *þá* ‘then’), henceforth ‘Group 1’ contexts, the subject is typically postfinite (subject-verb inversion), indicating V2, e.g. (1).

- (1) *þa cwæp he to him*
 then said he to them
 ‘then he said unto them...’ (BIHom-11:119.49.1511, van Kemenade 2012)

By contrast, in clauses where some other type of non-subject is clause-initial (e.g. a PP adjunct or object NP), henceforth ‘Group 2’ contexts, there is a split by subject type. Lexical (i.e. non-pronominal) subjects are typically postfinite, e.g. (2-a), but pronominal subjects are typically prefinite, reflecting an SVO system, e.g. (2-b).

- (2) a. [On twam þingum] hæfde **God** þæs mannes sawle gegodod
 in two things had God the man’s soul endowed
 ‘With two things God had endowed man’s soul’
 (ÆCHom I, 1.20.1, van Kemenade 2012)
- b. [Be ðæm] [**we**] magon suiðe swutule oncnawan ðæt...
 by that, we may very clearly perceive that
 ‘By that, we may perceive very clearly that ...’
 (CP 26.181.16, van Kemenade 2012)

Throughout ME, the subject-verb inversion pattern decreases in frequency, as subjects overall become increasingly prefinite. Various factors have been connected with this. Firstly, the Group 1/Group 2 distinction is said to remain relevant: in Group 1 contexts, subject-verb inversion persists (‘residual V2’ in Present-day English, Rizzi 1996), while in Group 2 contexts inversion is gradually lost. Specifically, subjects with certain properties (lexical, discourse-new) – which were typically postfinite in OE/early ME – become increasingly prefinite, in line with prenominal and discourse-given subjects (Haeberli 2002, van Kemenade and Westergaard 2012). Dialect has also been shown to interact with rates of subject-verb inversion in ME (Kroch and Taylor 1997, Kroch et al. 2000), specifically that in certain Northern texts both lexical and pronominal subjects invert across the board, indicating a more generalised V2 syntax compared to other regional varieties.

Using data from the Penn-Helsinki Parsed Corpus of Middle English (PPCME2, Kroch and Taylor 2000), extracted via CorpusSearch queries (Randall 2005), we examine how subject position (prefinite/postfinite) interacts over time with the various features suggested in the literature:¹

- subject type: pronominal/lexical
- subject’s information-structural status: given/new²
- clause-initial constituent:

¹We restrict our investigation to matrix clauses containing a finite verb and an overt subject.

²The PPCME2 does not annotate for information structure, so we used an approximation: as ‘given’ we took any subject which is pronominal or has overt definite marking; as ‘new’ we took any subject which is not pronominal and does not have overt definite marking.

- neg/discourse adverb (Group 1)
- PP/(non-discourse) adverbial/NP object (Group 2)
- dominant dialect of text: north/west-midlands/east-midlands/south³

While previous studies look at the interaction between subject position and one or two factors, HistoBankVis allows us to assess interactions between several factors at once. HistoBankVis provides exploratory and flexible access to complex data via three visualisation components: i) a compact matrix which allows the monitoring of data sparsity issues by providing an overview of the data across time periods, ii) difference histograms which grant access to differences between data distributions across time, and iii) dimension interactions based on the Parallel Sets technique (Bendix et al. 2005, Kosara et al. 2006), for exploring the interrelation between potentially interacting factors.

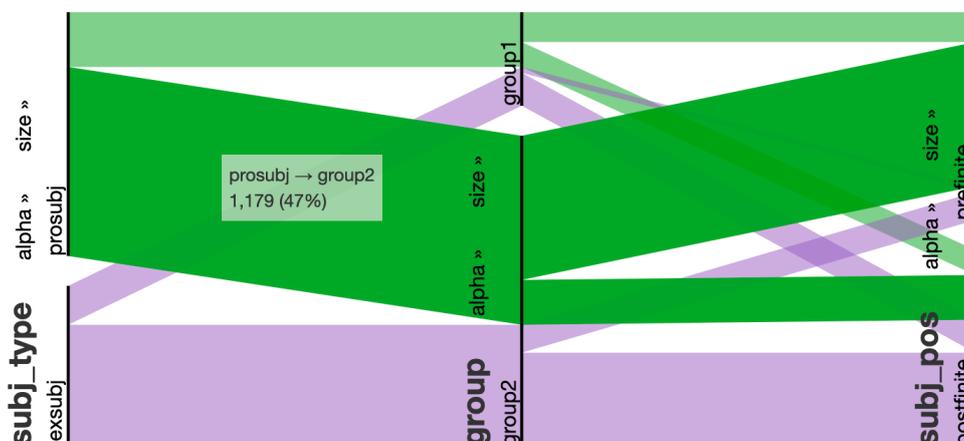


Figure 1: Dimension interaction between subject type (lexsubj/prosubj), clause-initial category (group1/group2) and subject position (prefinite/postfinite) from 1150-1250.

As a first step, we investigate previously suggested interactions via the dimension interaction component (i.e. parallel sets). Firstly, we look at the interaction between subject position, subject type and clause-initial category. The insights from HistoBankVis confirm that pronominal subjects lead the change towards becoming increasingly prefinite, with lexical subjects lagging behind. The previously claimed divergence between Group 1 and Group 2, however, is less clear-cut than often suggested. Compared to Group 2, Group 1 contexts are indeed conservative with respect to the increasing preference for prefinite subjects. Yet the dimension interactions show that Group 1 contexts are far from static: in the earliest period (1150-1250), see Figure 1, more than half of the pronominal subjects are already prefinite. Moreover, we find evidence that lexical subjects in Group 1 also follow suit by at least the third period (1350-1420). We suggest that these more nuanced findings for Group 1 are revealed through our decision to exclude *wh*-questions from the group. Furthermore, this change coincides with a striking decrease in clause-initial negation (neg), see the difference histogram in Figure 2.

³We follow the dialect classification on the corpus website: <https://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-4/info/texts-by-dialect.html>

This suggests future investigations should focus on clause-initial discourse adverb (da) contexts for further insights.

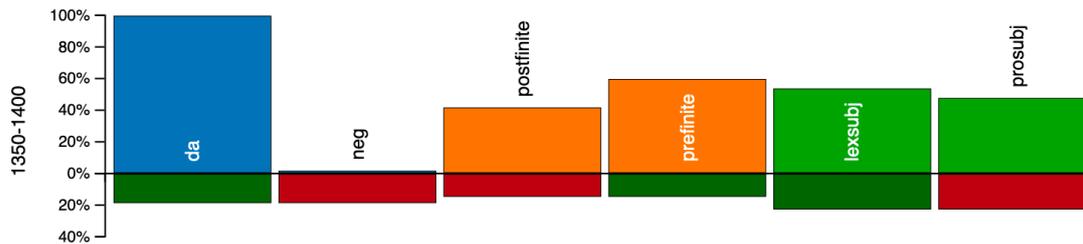


Figure 2: Difference histogram for Group 1 categories (da/neg), subject position and subject type from 1350-1420.

With respect to information structure, the previously suggested correlation between postfinite position and discourse-new subjects, and prefinite position and discourse-given subjects, is borne out overall. Moreover, this correlation weakens over time: by the last period (1420-1500) there is little information-structural effect on position. Since HistoBankVis allows us to investigate multiple factors at once, we can take the investigation further, examining whether information structure is in fact the driving force behind the differences between Group 1 and Group 2. It is difficult to isolate information structure from subject type as a factor, since given subjects will often be pronominal, and new subjects often lexical. Nevertheless, adding both factors into the dimension interaction indicates that subject type is more important than information structure, since the subject type correlation with subject position seems to be stronger overall.

Finally, we add in dialect as a factor, to examine the claim that Northern texts are particularly conservative to the change. Data sparsity is an issue here, with only one Northern text unambiguously dated (*Northern Prose Rule of St. Benet*, 1350-1420). This text has been shown to exhibit a generalised V2 system, with high levels of subject-verb inversion (Kroch and Taylor 1997, Kroch et al. 2000). However, when we include the other Northern texts, which are less clearly dated, the picture is much more variable, with higher rates of prefinite, in particular pronominal, subjects. The quick and easy comparison provided by HistoBankVis thus indicates that the trends across Northern texts from the period are particularly nuanced, and merit further research.

References

- Bendix, F., Kosara, R., and Hauser, H. (2005). Parallel sets: Visual analysis of categorical data. In *IEEE Symposium on Information Visualization INFOVIS*, pages 133–140. IEEE.
- Haeberli, E. (2002). Observations on the loss of verb-second in the history of English. In Zwart, C. and Abraham, W., editors, *Studies in Comparative Germanic Syntax*, volume 53, page 245. John Benjamins, Amsterdam.

- Hilpert, M. and Gries, S. T. (2016). Quantitative approaches to diachronic corpus linguistics. In Kytö, M. and Pahta, P., editors, *The Cambridge handbook of English historical linguistics*, pages 36–53. Cambridge University Press, Cambridge.
- Holmberg, A. (2015). Verb second. In Kiss, T. and Alexiadou, A., editors, *Syntax. Theory and Analysis: An International Handbook*, pages 342–383. de Gruyter, Berlin.
- Keenan, E. (1994). Creating anaphors: An historical study of the English reflexive pronouns. Ms. UCLA.
- Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., and Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In Kerren, A., Stasko, J. T., Fekete, J.-D., and North, C., editors, *Information Visualization*, pages 154–175. Springer, Berlin.
- Kosara, R., Bendix, F., and Hauser, H. (2006). Parallel Sets: interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):558–568.
- Kroch, A. and Taylor, A. (2000). The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania. Second edition.
- Kroch, A., Taylor, A., and Ringe, D. (2000). The Middle English verb-second constraint: A case study in language contact and language change. In Herring, S. C. and van Riemsdijk, P., editors, *Textual Parameters in Older Language*, pages 353–391.
- Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1(3):199–244.
- Kroch, A. S. and Taylor, A. (1997). Verb movement in Old and Middle English: Dialect variation and language contact. In van Kemenade, A. and Vincent, N., editors, *Parameters of Morphosyntactic Change*. Cambridge University Press, Cambridge.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3):273–309.
- Lightfoot, D. W. (2013). Types of explanation in history. *Language*, 89(4):e18–e38.
- Longobardi, G. (2001). Formal syntax, diachronic minimalism, and etymology: the history of French *chez*. *Linguistic Inquiry*, 32(2):275–302.
- Los, B. (2009). The consequences of the loss of verb-second in English: information structure and syntax in interaction. *English Language and Linguistics*, 13(1):97–125.
- Malkiel, Y. (1967). Multiple versus simple causation in linguistic change. In *To honor Roman Jakobson: Essays on the occasion of his seventieth birthday*, pages 1228–1246, The Hague. Mouton.

- Pintzuk, S. (1999). *Phrase structures in competition: Variation and change in Old English word order*. Garland, New York.
- Pintzuk, S., Taylor, A., and Warner, A. (2017). Corpora and quantitative methods. In Ledgeway, A. and Roberts, I., editors, *The Cambridge handbook of historical syntax*, pages 218–240. Cambridge University Press, Cambridge.
- Randall, B. (2005). *CorpusSearch2 User's Guide*. Philadelphia: Dept. of Linguistics, University of Pennsylvania. <http://corpussearch.sourceforge.net>.
- Rizzi, L. (1996). Residual verb second and the *wh*-criterion. In Belletti, A. and Rizzi, L., editors, *Parameters and functional heads: essays in comparative syntax*, pages 63–90. Oxford University Press Oxford, Oxford.
- Schätzle, C., Dennig, F. L., Blumenschein, M., Keim, D. A., and Butt, M. (2019). Visualizing linguistic change as dimension interactions. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 272–278, Florence, Italy. Association for Computational Linguistics.
- Schätzle, C., Hund, M., Dennig, F. L., Butt, M., and Keim, D. A. (2017). HistoBankVis: Detecting language change via data visualization. In Bouma, G. and Asedam, Y., editors, *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, pages 32–39, Linköping. Linköping University Electronic Press.
- van de Velde, F., De Smet, H., and Ghesquière, L. (2013). On multiple source constructions in language change. *Studies in Language*, 37(3):473–489.
- van Kemenade, A. (1987). *Syntactic case and morphological case in the history of English*. Foris, Dordrecht.
- van Kemenade, A. (2012). Rethinking the loss of verb second. In Nevalainen, T. and Traugott, E. C., editors, *The Oxford Handbook of the History of English*, pages 823–834. Oxford University Press, Oxford.
- van Kemenade, A. and Westergaard, M. (2012). Verb-Second Variation in Middle English. In Meurman-Solin, A., Lopez-Couso, M. J., and Los, B., editors, *Information structure and syntactic change in the history of English*, pages 87–118. Oxford University Press, Oxford.
- Weinreich, U., Labov, W., and Herzog, M. I. (1968). Empirical foundations for a theory of language change. In Lehmann, W. P. and Malkiel, Y., editors, *Directions for Historical Linguistics*, pages 95–195. University of Texas Press, Austin, TX.