

Paradigmatic change and redistribution of functional load: The case of relative clauses in scientific English

Marie-Pauline Krielke, Stefania Degaetano Ortlieb, Katrin Menzel, Elke Teich
Saarland University, Germany

We present a study of diachronic change in language use from the perspective of lexicogrammatical paradigms in the period of Late Modern English. While Late Modern English does not count as a particularly change-intensive period of the English language, it is that period in which many of our modern registers and genres evolved --- including the language of science (cf. Halliday & Martin, 1993; Mair & Hundt, 1999). Focusing on scientific English, one relevant lexicogrammatical paradigm is that of *wh*-relativizers (*whereby*, *whereon*, *wherein* etc.; see example 1), which exhibits a high productivity in early scientific writing (late 17th / early 18th centuries) and becomes reduced to just a few items later on with a strong preference for *which*. The question arises whether analytic *which* (preposition + *which*, see example 2) takes over the functions associated with the earlier, synthetic relativizers or whether some of the functional load is redistributed to other areas of grammar and if so, to which.

- (1) [...] he propounds a Method, *whereby* the Solution of such may be found [...]. (RSC, unknown, 1674)
- (2) The air-engine described in this paper consists of a pump *by which* air is compressed into a heated receiver [...]. (RSC, J. P. Joule, 1850)

Informed by systemic-functional grammar (Halliday, 1985), according to which language is best characterized by use in context, we explore the changes affecting the *wh*-relativizer paradigm as shown in scientific writing. As a data set we use the Royal Society Corpus (RSC, Kermes et al., 2016) v.4.0.¹ To explore paradigmatic change, we employ a distributional semantics approach known as *word embedding* (Word2Vec; cf. Mikolov et al., 2013). Based on a visualization of the RSC (Fankhauser & Kupietz, 2017, Fankhauser & Kupietz, to appear), we can see how paradigms of words change over time (see Figure 1).

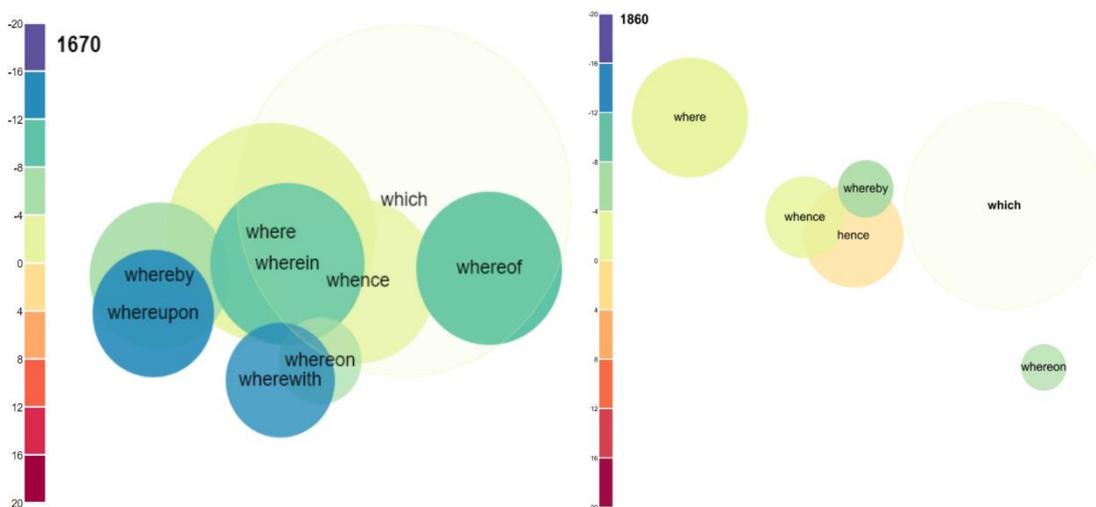


Figure 1. Wh-relativizers in two decades: 1670s (left) vs. 1860s (right)

¹ The RSC is an open source corpus available at http://fedora.clarin-d.uni-saarland.de/rsc_v4/, containing journal publications of the Philosophical Transactions and Proceedings of the Royal Society of London and spanning from 1665 to 1869 (~ 32 million tokens and ~ 10,000 documents).

The color scheme refers to frequency change: from violet, indicating decreasing frequency, to red, indicating increasing frequency. The size of the bubbles indicates relative frequency in the selected decade (ibid.). On this basis, we can observe not only shrinking paradigms but expanding paradigms, too. *ing*-forms are one such paradigm (see Figure 2)

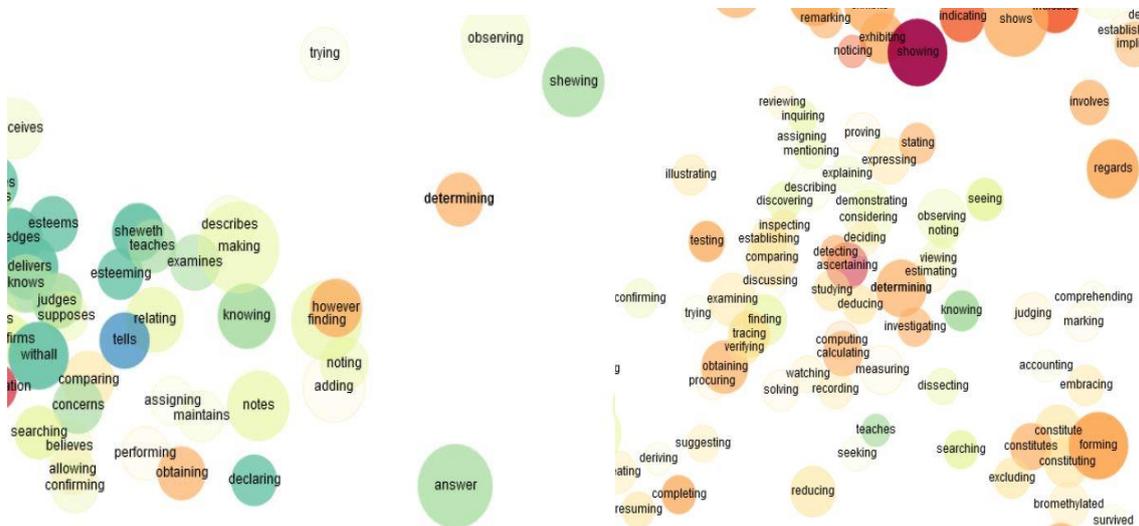


Figure 2. *ing*-forms in two decades: 1670s (left) vs. 1860s (right)

For an example from the RSC see (3) below.

- (3) Two galvanometers were used; one by Ruhmkorff, *containing* fine wire, and very delicate in its action. (RSC, Earl of Rosse, 1850)

First, to confirm what we observe in the word embedding regarding the *wh*-relativizers (Figure 1), we diachronically compare the distributions of all synthetic vs. analytic *wh*-relativizers (e.g. *whereby* vs. *by which*) by extracting the synthetic ones as well as bigrams consisting of a preposition followed by *which* per periods of 50 years. As *ing*-forms can occur in non-finite relative clauses, they too may present an alternative to *wh*-relativizers. Considering the reduction in *wh*-relativizers and the increase in *ing*-forms over time, a possible explanation could be that *ing*-forms act as a substitute for obsolete *wh*-relativizers, leading to a more condensed way of expression typical of scientific language (Halliday and Martin, 1993). By applying pointwise relative entropy or Kullback-Leibler divergence (KLD; cf. Fankhauser et al., 2014; Degaetano-Ortlieb & Teich, 2018) we obtain the most distinctive alternative (i.e. *whereby* vs. *by which* vs. *ing*-form) across time periods.

Our results show that synthetic forms (types) do indeed decline over time and most of them disappear. However, analytic *which* stays stable over time, except for *in which* showing a steep increase. Further inspection of the syntagmatic context shows that *in which* increasingly appears with nouns denoting manner, such as *the manner in which* or *the mode in which*. Furthermore, pointwise KLD shows that synthetic *wh*-relativizers are clearly typical of the early periods (second half of 17th c.), while *ing*-forms and analytic *which* are distinctive for the later periods (mid 19th c.). Together, these results indicate a redistribution of functional load from synthetic *wh*-relativizers to analytic *which*, on the one hand, and non-finite relative clauses with *ing*-forms, on the other hand.

References

- Degaetano-Ortlieb, S. & Teich, E. (2018) Using relative entropy for detection and analysis of periods of diachronic linguistic change. In Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, COLING 2018, pages 22–33, Santa Fe, NM, USA.
- Fankhauser, P., Knappen, J. & Teich, E. (2014) Exploring and Visualizing Variation in Language Resources. In Proceedings of the 9th LREC, pages 4125–4128, Reykjavik, Iceland.
- Fankhauser, P. & Kupietz, M. (2017) Visualizing Language Change in a Corpus of Contemporary German. In Proceedings of the Corpus Linguistics International Conference, Birmingham, UK.
- Fankhauser, P. & Kupietz, M. (To appear) Analyzing domain specific word embeddings for a large corpus of contemporary German. In Proceedings of the Corpus Linguistics International Conference 2019, Cardiff, UK.
- Halliday, M.A.K. & Martin, J.R. (1993) *Writing Science: Literacy and discursive power*. The Falmer Press, London.
- Halliday, M.A.K. (1985) *An Introduction to Functional Grammar*. Arnold, London.
- Kermes, H., Degaetano-Ortlieb, S., Khamis, A., Knappen, J. & Teich, E. (2016) The Royal Society Corpus: From uncharted data to corpus. In Proceedings of the 10th LREC, Portoroz, Slovenia.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. & Dean, J. (2013) Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems. 3111–3119.
- Mair, C. & Hundt, M. (1999) “Agile” and “uptight” genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics*, 4(2), 221–242.