

## A ‘hands-off’ approach to modelling constructional change with word embeddings: current challenges and future prospects

Lauren Fonteyn & Enrique Manjavacas (Leiden University)

For a long time, the construction grammarian’s interest in the functional-semantic process accompanying or underlying the formal side of constructional change seemed difficult to reconcile with their increasing reliance on quantitative corpus methodologies. This has, however, started to change in recent years: following the rise of an increasingly powerful body of computational models that aim to approximate the meaning of words over time by numerically encoding their linguistic context (or ‘distributional properties’), the path towards a fully data-driven, quantifiable and measurable approach to functional-semantic change appears to have been cleared (e.g. Sagi, Kaufmann & Clark 2011; Hamilton, Leskovec & Jurafsky 2016a; Hamilton, Leskovec & Jurafsky 2016b; Dubossarsky et al. 2019). At present, these models have mainly been applied in more coarse-grained studies to capture and quantify some aspect of lexical semantic change (and, occasionally, grammaticalization; see Luo, Jurafsky & Levin 2019), but they have found their way into historical linguistic case studies of constructional change too (e.g. Perek 2016; Budts & Petré 2020).

The first aim of our contribution (Fonteyn & Manjavacas 2021) is to rely on models and methods developed in the computational literature on semantic change to set out a fully ‘hands-off’ computational procedure to study constructional change. The specific case we address concerns the development of the phrasal expression *to death* from a literal, resultative phrase (1) into an ‘amplifying’ expression (2) (as previously described in Hoeksema & Jo Napoli 2008; Claridge 2011; Margerie 2011; Blanco Suárez 2017).

- (1) “because she durst compare hirself, with Diana was by the goddesse wounded to death with an arrow;” (1641, EEBO)
- (2) “Who ever heard or imagined grandparents not being just pleased to death at the coming of their first grandchild!” (1930, COHA)

Following the research design of Perek (2016, 2018), we deploy hierarchical clustering algorithms over distributed meaning representations (obtained through word2vec; Mikolov et al. 2013) in order to capture the evolution of the semantic space of verbs types that collocate with *to death*. We then describe the diachronic trends by means of a monotonic effects regression model, which helps us to provide an accurate picture of the underlying diachronic process. The methodology we outline may help tackle some common challenges in the use of functional-semantic vector representations to study similar cases of constructional change.

The second aim of this talk is to pinpoint some (remaining) challenges for vector-based approaches to modelling constructional change. First, we briefly summarize what we consider to be the most important challenges for computational analyses of constructional change, including the open-ended question on how to strike the right balance between robustness and transparency, and between automation and manual/introspective correction. Finally, we highlight how the ‘lumped’ vector representations generated by type-based models may introduce problems into fully hands-off procedures, and introduce the token-based models we are currently developing as part of the MacBERTH infrastructure project.

Blanco Suárez, Zeltia. 2017. Death-related intensifiers in the history of the English language: Grammaticalisation and other processes of language change. Spain: Universidade de Santiago De Compostela PhD dissertation.

- Budts, Sara & Peter Petré. 2020. Putting connections centre stage in diachronic construction grammar. In Lotte Sommerer & Elena Smirnova (eds.), *Nodes and Networks in Diachronic Construction Grammar*, 317–352. Amsterdam: John Benjamins.
- Claridge, Claudia. 2011. *Hyperbole in English: A Corpus-based Study of Exaggeration*. Cambridge: Cambridge University Press.
- Dubossarsky, Haim, Simon Hengchen, Nina Tahmasebi & Dominik Schlechtweg. 2019. Time-Out: Temporal Referencing for Robust Modeling of Lexical Semantic Change. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 457–470. Florence, Italy: Association for Computational Linguistics.
- Fonteyn, Lauren & Enrique Manjavacas. 2021 (in press). Adjusting scope: a computational approach to case-driven research on semantic change. *Computational Humanities Research*.
- Hamilton, William L., Jure Leskovec & Dan Jurafsky. 2016a. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1489–1501. Berlin, Germany: Association for Computational Linguistics. <https://doi.org/10.18653/v1/P16-1141>.
- Hamilton, William L., Jure Leskovec & Dan Jurafsky. 2016b. Cultural shift or linguistic drift? Comparing two computational measures of semantic change. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2116–2121. Austin, Texas: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D16-1229>.
- Hoeksema, Jack & Donna Jo Napoli. 2008. Just for the hell of it: A comparison of two taboo-term constructions. *Journal of Linguistics* 44(2). 347–378. <https://doi.org/10.1017/S002222670800515X>.
- Luo, Yiwei, Dan Jurafsky & Beth Levin. 2019. From Insanely Jealous to Insanely Delicious: Computational Models for the Semantic Bleaching of English Intensifiers. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, 1–13. Florence, Italy: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W19-4701>.
- Margerie, H el ene. 2011. Grammaticalising constructions: to death as a peripheral degree modifier. *Folia Linguistica Historica* 45(Historica vol. 32). <https://doi.org/10.1515/flih.2011.005>.
- Mikolov, Tom as, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Yoshua Bengio & Yann LeCun (eds.), 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*. <http://arxiv.org/abs/1301.3781>.
- Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1). <https://doi.org/10.1515/ling-2015-0043>.
- Perek, Florent. 2018. Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1). 65–97. <https://doi.org/10.1515/cllt-2016-0014>.
- Sagi, Eyal, Stefan Kaufmann & Brady Clark. 2011. Tracing semantic change with Latent Semantic Analysis. In *Kathryn Allan & Justyna A. Robinson (eds.), Current Methods in Historical Semantics*. Berlin, Boston: DE GRUYTER. <https://doi.org/10.1515/9783110252903.161>.