Towards a communicative explanation of diachronic language change: an information-theoretic approach

Stefania Degaetano-Ortlieb and Elke Teich, Saarland University
Keywords: diachronic language change, information theory, rational communication, scientific writing

There is ample descriptive evidence of different patterns of diachronic language change but explaining change, including actuation (Weinreich et al. 1968), remains a challenge. While we cannot predict change sparked by external events (e.g. boosts of new vocabulary), we can however characterise the linguistic dynamics of change and seek language-internal explanations. For example, adopting a frequency-based framework, Bybee (2010) speaks of an innovation-habituation-grammaticalization cycle or De Smet (2016) observes that conventionalization is a prerequisite for innovative uses against the background of cognitive grammar, thus providing partial explanations of change.

We here suggest adopting the perspective of rational communication on language change as a linking theory. Under this perspective, we acknowledge the nature of language as a dynamic, adaptive system that, while continuously changing, strives to maintain its communicative function. To provide a link between modelling and explaining change, we propose an information-theoretic framework, according to which the amount of information needed for encoding a linguistic unit reflects its *predictability in context* – commonly known as *surprisal* – and, importantly, is proportional to cognitive effort (Hale 2001, Levy 2008, Crocker et al. 2016). A related concept, also indexing processing effort, is *entropy* which denotes the *degree of uncertainty* about upcoming material in on-line language processing. Low surprisal and entropy ease processing and are therefore said to be communicatively beneficial.

We will show how we use state-of-the-art computational language models (n-gram models, embeddings) to estimate surprisal and entropy diachronically and detect linguistic features involved in change as well as periods and cycles of change. Importantly, we link this to explanation by hypothesising that for a specific change to occur and be persistent, it should be beneficial for communication, i.e. reduce or at least maintain (average) surprisal and/or entropy. Also, we will show that the information-theoretic perspective allows us to categorize our empirical findings into two interacting diachronic mechanisms, conventionalization and diversification, which links up well with previous works such as the works mentioned above by De Smet and Bybee.

Our focus is on scientific English, which provides us with a good control over discourse domain and easy linkage with external historical events (e.g. the emergence of modern chemistry). We investigate the overall hypothesis of communicative optimization, stating that scientific English developed an optimal code for expert-to-expert communication over time. We will illustrate our findings with data from the late Modern period based on the Royal Society Corpus (Fischer et al. 2020). Across different computational models, we observe the same trend of temporary peaks of innovative language use (surprisal increase) and an overall decreasing set of linguistic choices (entropy reduction). Our approach thus allows us to observe in one model the effects of external factors involved in language change (e.g. temporary vocabulary expansion in the field of chemistry indexed by high entropy) and internal constraints, such as convergence on the use of specific linguistic forms in particular contexts (e.g. *ing*-forms in postmodifier function indexed by low entropy).

References

Bybee, J. L. (2010). *Language, Usage and Cognition*, Cambridge University Press, Cambridge.

Crocker, M. W., Demberg, V., and Teich, E. (2016). Information density and linguistic encoding (IDeaL). *KI-Künstliche Intelligenz*, *30*(1), 77-81.

De Smet, H. (2016). How gradual change progresses: The interaction between convention and innovation. *Language Variation and Change*, 28:83–102.

Fischer, S., Knappen, J., Menzel, K., and Teich, E. (2020). The Royal Society Corpus 6.0: Providing 300+ years of scientific writing for humanistic study. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 794–802, Marseille, France. ELRA.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd meeting of the North American Chapter of the Association for Computational Linguistics*, pages 1–8, Stroudsburg, PA, USA.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177.

Weinreich, U, Labov W. and Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W.P. Lehmann and Y. Malkiel (eds.), *Directions for Historical Linguistics*. University of Texas Press, Austin, Texas, pages 95-195.