

THE EVOLUTIONARY DYNAMICS OF LANGUAGE AS A FUNCTION OF DEMOGRAPHY

DIRK PIJPOPS^{1,2}, KATRIEN BEULS³, and FREEK VAN DE VELDE^{*2}

*Corresponding Author: freek.vandavelde@kuleuven.be

¹Research Foundation Flanders (FWO), Brussels, Belgium

²QLVL, University of Leuven, Leuven, Belgium

³AI Laboratory, Vrije Universiteit Brussel, Brussels, Belgium

1. Introduction

Lieberman et al. (2007) aimed to quantify the evolutionary dynamics of language by investigating the rise of the English regular past tense inflection, which they equated with the weak *-ed* suffix. Yet, their bold conclusion that “*the half-life of an irregular verb scales as the square root of its usage frequency: a verb that is 100 times less frequent regularizes 10 times as fast*” (Lieberman et al., 2007, p.713) has successively attracted criticism from scholars in the fields of historical and evolutionary linguistics. First, Carroll, Svare, & Salmons (2012) showed that this constant regularization rate does not hold true for the closely-related German language. Second, Cuskley et al. (2014) found that the rise of the English weak *-ed* suffix is not driven by forces endogenous to language, such as analogy, but rather by external forces, such as new verbs entering the language through language contact.

We will reassess the constant-rate controversy by (i) extending the methodological scope with agent-based modeling, and (ii) extending the number of languages going beyond the German-English distinction, adding Dutch.

Our results show that the constant rate does not hold. If language change is co-determined by external forces resulting in languages adapting to its niche (Lupyan & Dale 2016) this is exactly what one would expect, since English, Dutch and German have endured external pressures to a different degree. We will focus on the influence of demographic change. In particular, we investigate the growth of cities and the resulting koineization due to migration in the three

language areas since the Middle Ages. The three different degrees of urbanization have led to different degrees of dialect contact, which could in turn, as we will argue, lead to different regularization rates. To support this claim, we will present both empirical evidence from linguistic and demographic databases, as well as the results of a computational simulation.

2. Empirical data

2.1. Linguistic data

To obtain a clear picture of the linguistic situation, we included the data on English from Lieberman et al. (2007) and the data on German from Carroll et al. (2012), and complemented these with our own Dutch data. This enables us to track the development of the past tense system of these three languages over a 1000 year period (800-1800).

2.2. Demographic data

For the demographic data, we make use of the databases of Bairoch et al. (1988), De Vries (1984), and Mitchell (1998). In particular, we compare the population growth of the largest cities in the English, Dutch and German language areas in each particular time period from 800-1800. Historical research has shown that the exponential growth of urban population cannot be reduced to natural growth, but is driven by immigration as well, both of foreigners and of by a rural exodus from the larger agglomeration, leading to dialect contact. We then observed correlations between the success of the weak inflection and the amount of demographic upheaval.

3. Simulation

A correlation between a demographic and a linguistic trend does not automatically entail a causation between the former and the latter, however. To further substantiate our claim, we therefore turn to an agent-based computer simulation. In this simulation, agents store exemplars or tokens of what they hear (cf. Pijpops et al., 2015), rather than type states (cf. Colaïori et al., 2015), and use these to produce novel forms. We find that (i) the weak inflection does not require special status as the single regular inflection in order to explain the tendencies observed in reality; (ii) replacement of verbs can indeed cause a continued rise of the weak inflection, even after a stable equilibrium between weak and strong verbs has emerged, confirming Cuskley et al. (2014); and most importantly (iii) if our

current understanding of language, as implemented in the simulation, is correct, demography does indeed affect the rise of the weak inflection.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive feedback as well as our colleague Isabeau De Smet for her help in this study.

References

- Bairoch, P., Batou, J. & Chèvre, P. (1988). *La population des villes européennes de 800-1850*. Geneva: Librairie Droz.
- Carroll, R., Svare, R., & Salmons, J. (2012). Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics*, 2(2), 153–172.
- Colaïori, F., Castellano, C., Cuskley, C., Loreto, V., Pugliese, M., & Tria, F. (2015). General three-state model with biased population replacement: Analytical solution and application to language dynamics. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 91(1–1), 12808.
- Cuskley, C., Pugliese, M., Castellano, C., Colaïori, F., Loreto, V., & Tria, F. (2014). Internal and External Dynamics in Language: Evidence from Verb Regularity in a Historical Corpus of English. *Plos One*, 9(8), e102882.
- De Vries, J. 1984. *European Urbanization 1500-1800*. Cambridge, Mass.: Harvard University Press.
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449(7163), 713–716.
- Lupyan, G. & Dale, R. 2016. Why are there different languages? The role of adaptations in linguistic diversity. *Trends in Cognitive Science*.
- Mitchell, B.R. 1998. *International Historical Statistics: Europe 1750-1993*. Springer.
- Pijpops, D., Beuls, K., & Van de Velde, F. (2015). The rise of the verbal weak inflection in Germanic. An agent-based model. *Computational Linguistics in the Netherlands Journal*, 5, 81-102.