LINGUISTIC AND NON-LINGUISTIC CORRELATES IN THE EVOLUTION OF PHONOTACTIC DIVERSITY

ANDREAS BAUMANN[*1], THERESA MATZINGER[1], and NIKOLAUS RITT[1]

[*]Corresponding Author: andreas.baumann@univie.ac.at
[1]Department of English and American Studies, University of Vienna, Vienna, Austria

Linguistic dynamics have been hypothesized to be driven by ecological factors such as population size or social structure (see Nettle, 2012 for an excellent overview). Particularly, there is an ongoing debate as to whether population size can be seen as an explanatory factor in the evolution of phonemic richness (Atkinson, 2011; Bybee, 2011; Hay & Bauer, 2007; Wichmann, Rama, & Holman, 2011; see also Moran, McCloy, & Wright, 2012 for critical discussion). In this regard, the evolution of larger sublexical constituents, i.e. sequences of sounds below the word level, has gained much less attention (but see Maddieson, 2013 or Rama, 2013). Moreover, studies on the connection between ecological factors and linguistic properties were primarily comparative in nature, although the parallel evolution of social structure and language in individual linguistic strands may also provide useful insights into the mechanics that drive language evolution (see Bybee, 2011; Pagel, Atkinson, & Meade, 2007; Trudgill, 2004).

In this paper, we conceptualize phonotactic items (sequences of sounds) as culturally transmitted pieces of linguistic knowledge, i.e. competence constituents in their own right, which spread through populations just like single sounds, words or constructions (Croft, 2000; Ritt, 2004). Phonotactic items should therefore be subject to similar evolutionary pressures and mechanisms. We investigate the diachronic development of diversity of the phonotactic inventory in the history of English from Middle English to Present Day English (using historical data from PPCME2, PPCEME, PPCMBE and COHA, and phonological transcriptions from ECCE and CMU). We focus on word final phonotactics because changes are most likely to occur at this prosodically weak position, and for methodological reasons (fully phonologically analyzed historical texts are not available for early periods). We find that the diversity of word-final coda phonotactics has been increasing through the past 800 years, and that the

DOI:10.12775/3991-1.003

evolution of phonotactic diversity is strongly related to that of network characteristics that can be derived from population size.

Our approach goes like this: for each period of 50 years from 1150 to 2000, we computed *true diversity* (cf. Tuomisto, 2010) based on the respective frequency distributions of word-final consonant sequences. The resulting trajectory indicates that English phonotactics became more diverse. We then retrieved trajectories for potentially related features that fall into three categories: (a) linguistic features (size of diphone inventory; consonant-inventory size; syntheticity; analyticity; cf. Szmrecsanyi, 2012), (b) socio-geographic features (population size; populated area; population density of the English speaking community; estimates taken from Wrigley & Schofield, 1981 and more recent census data), and (c) network features directly derived from population size under the assumption of a scale-free small-world network (network diameter; clustering coefficient; Barabási, 2016). In total, this amounts to ten trajectories.

In order to compare the trajectories to each other and to find out which development matches that of phonotactic diversity best, we use autocorrelation-driven time-series clustering (this has – in contrast to e.g. Pearson or Minkowski-distance based procedures – the advantage of also taking the temporal structure into account; see Montero & Vilar, 2014 and references therein). We find that the evolution of phonotactic diversity correlates most strongly with that of the computed clustering coefficient (albeit in a negative way: high clustering corresponding to low diversity) and with the trajectories of population density, populated area and network diameter (all positively correlated). Phonotactic inventory size (i.e. diversity as measured in Rama, 2013) correlates less strongly with factors in that group (which entails as a corollary that dynamics in phonotactic diversity are not just a reflex of increased lexical diversity due to loan import etc.). The remaining trajectories (notably population size together with the other linguistic features) form separate groups.

Our analysis yields a number of insights. First, it suggests that it is probably not population size itself (and associated exposure to drift effects) which directly affects linguistic evolution but rather more immediate (but related) factors that determine the amount and heterogeneity of linguistic interactions. Indeed, increased clustering (i.e. the tendency of forming small groups) has been shown to decrease growth of new variants (Miller, 2009). Likewise, high population density can be argued to promote the spread of linguistic constituents (even if they *a priori* have deficient reproductive properties like sequences of consonants in the prosodically weak coda-position). Second, on a more methodological level, we argue that language-dating methods which are based on phonotactic diversity

(Rama, 2013) must take population size and related factors into account in order to prevent the method from just reflecting world-wide population increase in the past centuries. Finally, in agreement with Bybee (2011), we stress that measures of linguistic diversity which also take token frequency into account (such as entropy or true diversity) might be more profitable for the research on language evolution than counts of types (e.g. phoneme or diphone inventory size, lexicon size).

Acknowledgements

We would like to thank Klaus Hofmann for helpful comments on this study. This research was supported by Austrian Science Fund (FWF, grant No. P27592-G18).

References

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, *322*(6027), 346–349.

Barabási, A.-L. (2016). *Network science*. Cambridge: Cambridge University Press.

Bybee, J. (2011). How plausible is the hypothesis that population size and dispersal are related to phoneme inventory size? Introducing and commenting on a debate. *Linguistic Typology*, *15*, 147–153. https://doi.org/10.1515/LITY.2011.009

Croft, W. (2000). *Explaining language change: An evolutionary approach*. *Longman linguistics library*. Harlow, England, New York: Longman.

Hay, J., & Bauer, L. (2007). Phoneme inventory size and population size. *Language*, *83*(2), 388–400.

Maddieson, I. (2013). Syllable Structure. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Max Planck Digital Library. Retrieved from http://wals.info/chapter/12

Miller, J. (2009). Spread of infectious disease through clustered populations. *J R Soc Interface*, *6*, 1121–1134.

Montero, P., & Vilar, J. (2014). TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software*, *62*(1).

Moran, S., McCloy, D., & Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language*, *88*(4), 877–893. https://doi.org/10.1353/lan.2012.0087

Nettle, D. (2012). Social scale and structural complexity in human languages. *Phil. Trans. R. Soc. B*, *367*, 1829–1836.

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature, 449*(7163), 717–720.

Rama, T. (2013). Phonotactic diversity predicts the time depth of the world's language families. *PloS one.* Advance online publication. https://doi.org/10.1371/journal.pone.0063238

Ritt, N. (2004). *Selfish sounds and linguistic evolution: A Darwinian approach to language change*. Cambridge: Cambridge University Press.

Szmrecsanyi, B. (2012). Analyticity and syntheticity in the history of English. In T. Nevalainen & E. C. Traugott (Eds.), *The Oxford handbook of the history of English* (pp. 654–665). Oxford: Oxford University Press.

Trudgill, P. (2004). Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology, 8*(3), 305–320.

Tuomisto, H. (2010). A consistent terminology for quantifying species diversity? Yes, it does exist. *Oecologia, 164*, 53–860. https://doi.org/10.1007/s00442-010-1812-0

Wichmann, S., Rama, T., & Holman, E. (2011). Phonological diversity, word length, and population sizes across languages: The ASJP evidence. *Linguistic Typology, 15*(2), 177–197.

Wrigley, E. A., & Schofield, R. (1981). *The Population History of England, 1541–1871. A reconstruction*. Harvard: Harvard University Press.