Causality in historical language change:
The co-evolution of population structure and lexical diversity

Christian Bentz

Dept. of Theoretical and Applied Linguistics, University of Cambridge

Darwin College, Silver Street, CB3 9EU, UK

chris@christianbentz.de

Theoretical background

In studies of historical language change causality is often devided into two domains: *internal* and *external* causes of change. In this view languages are generally closed systems that inherit structural features from their proto-languages and follow 'natural' or 'normal' pathways of 'internal' change, i.e. general principles of *assimilation*, *analogical extension* and *analogical leveling* (see [6,7,11] for discussion). Language 'external' causes then refer to language contact, i.e. child bilingualism or adult second language learning.

However, within a language as *Complex Adaptive System* (CAS) account [1,4] linguistic structures emerge from the communicative interactions of speakers and hearers. Based on this framework, the theoretical part of this presentation suggests that a) the internal/external distinction of causes of change is misleading and b) that causes of change have to be sought in the make-up of speaker populations, i.e. the accumulation of differing language learning abilities. First and second language acquisition are paradigm examples of such differences that can have repercussions on language usage and change [9,12,13].

Statistical modeling

Quantitative studies corroborated the language as CAS argument by correlating measures of linguistic structure with measures of population structure and language contact [2,3,8]. These studies relied on the *World Atlas of Language Structures* [5] as a source of linguistically meaningful response variables for statistical modeling. However, the WALS often reverts to crude categorizations of linguistic features.

As an alternative, the methodological part of this presentation discusses lexical diversity as meaningful linguistic concept. Lexical diversity relates to word frequency distributions and is measured based on indices used in studies of biodiversity and in quantitative linguistics, i.e. Zipf-Mandelbrot's law [14], Shannon entropy [10] and type-token ratios. To estimate the cross-linguistic range of lexical diversities, parallel translations of the *Universal Declaration of Human Rights* into 335 languages of ca. 50 different families are used. Information about population structure is taken from a data set that comprises 226 languages for which native and non-native speaker numbers are given. To link lexical diversities with non-native speaker ratios, three types of statistical model are discussed: simple linear regression, linear mixed-effects and phylogenetic regression.

Though the statistical mantra of 'correlation is not causation' still holds, this presentation argues that qualitative and quantitative evidence can converge to elicit some of the causes of language change and diversity.

References

[1] Clay Beckner, Nick C Ellis, Richard Blythe, John Holland, Joan Bybee, Morten H Christiansen, Diane Larsen-Freeman, William Croft, and Tom Schoenemann. Language is a complex adaptive system. *Language Learning*, 59(December):1–26, 2009.

[2] Christian Bentz and Bodo Winter. The impact of L2 speakers on the evolution of case marking. In Thomas C. Scott-Phillips, Mónica Tamariz, Erica A Cartmill, and James R Hurford, editors, *The evolution of language. Proceedings of the 9th international conference (EVOLANG9)*, pages 58–64, Singapore, 2012. World Scientific.

[3] Christian Bentz and Bodo Winter. Languages with more second language speakers tend to lose nominal case. *Language Dynamics and Change*, 3:1–27, 2013.

[4] Morten H Christiansen and Nick Chater. Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(5):489–509, 2008.

[5] Matthew S. Dryer and Martin Haspelmath, editors. *World atlas of language structures online*. Max Planck Digital Library, Munich, 2013.

[6] Mari C. Jones and Ishtla Singh. *Exploring language change*. Routledge, New York, 2005.

[7] David W. Lightfoot. *Principles of diachronic syntax*. Cambridge University Press, Cambridge, 1979.

[8] Gary Lupyan and Rick Dale. Language Structure Is Partly Determined by Social Structure. *PloS ONE*, 5(1):e8559, January 2010.

[9] John H McWhorter. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford University Press, New York, 2007.

[10] Claude E. Shannon and Warren Weaver. *The mathematical theory of communication*. The University of Illinois Press, Urbana, 1949.

[11] Sarah Grey Thomason and Terrence Kaufman. *Language contact, creolization, and genetic linguistics*. University of California Press, Berkeley, Los Angeles, Oxford, 1988.

[12] Peter Trudgill. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press, Oxford, 2011.

[13] Alison Wray and George W Grace. The consequences of talking to strangers : Evolutionary corollaries of socio-cultural influences on linguistic form. 117:543–578, 2007.

[14] George Kingsley Zipf. *Human behavior and the principle of least effort*. Addison-Wesley, Cambridge (Massachusetts), 1949.