PNAS (May 9, 2017) 114[19]: 4851–4853, doi: www.pnas.org/cgi/doi/10.1073/pnas.1703509114

COMMENTARY: Modeling language shift

Anne Kandler a, 1, and James Steele b,c

[a] Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany; and [b] University College London, London WC1E 6BT, United Kingdom; and [c] Department of Archaeology, School of Geography, Archaeology and Environmental Sciences, University of the Witwatersrand, WITS 2050, South Africa

Author contributions: A.K. and J.S. wrote the paper.

The authors declare no conflict of interest.

See companion article 10.1073/pnas.1617252114.

[1]To whom correspondence should be addressed. Email: anne_kandler@eva.mpg.de.

It is estimated that there are as many as 6,000 distinct languages currently spoken, but this cultural diversity is rapidly disappearing. In terms of the fraction of the total being lost, the estimated current rate of language extinction exceeds the rate of loss of biodiversity (1–3). Small geographical ranges and small speaker populations, especially in parts of the world experiencing high economic growth, are among the strongest indicators of language extinction risk (4). Most of the recent language extinction events are caused by language shift rather than the extinction of the population speaking this language (5). Language shift is the process whereby members of a community in which more than one language is spoken abandon their original vernacular language in favor of another. Especially in language contact situations, people are confronted with choices about which language to speak. Processes of globalization, urbanization, and long-distance economic migration, have led to increased interactions between groups speaking different languages, and therefore to a need of a common language of communication. Historically, in situations of warfare, disaster, or repression, the dominant language may be forcibly imposed as a deliberate policy of its native speakers. In other cases, inequality may be expressed less directly. Often the language seen as more modern, useful, or giving access to greater social mobility and economic opportunities is chosen as the lingua franca, thereby driving the process of shift (6).

Research into language shift has been conducted in a variety of scientific disciplines, from linguistics and sociolinguistics to economics and physics. In one approach, language shift is modeled as a competition between linguistic communities for speakers, analogous to the competition dynamics between interacting species in ecology (but with speakers capable of “switching species”). These models make the simplifying assumption that languages are fixed entities, and that only speaker frequencies in the population can vary. Prochazka and Vogl (7) contribute to this literature by developing a cellular automata framework to model language shift on a fine spatial scale. Linguists recognize that frequency of use, proficiency, and perceived value are among the most important immediate factors affecting persistence or endangerment of minority languages (8), and Prochazka and Vogl’s (7) model assumes that such factors are captured by a measure of the local abundance of speakers. The authors show that neighborhood concentrations of speakers of the minority language are good predictors of its persistence (or of a slowing of its rate of decline).

Modeling enables us to gain a better understanding of the process of language shift, by identifying demographic, socioeconomic, cultural, and linguistic processes that can generate spatial and temporal patterns similar to those seen empirically. The comparison of theory with data is one of the crucial components of this literature, although this does not remove the problem of equifinality (especially where the data are sparse). Subsequent formal analysis of the models can also shed light on possible future outcomes of language shift. Additionally, models can be used as an artificial experiment. In the past, serious concerns over the loss of linguistic diversity have driven governments and international organizations to actively engage in the maintenance of endangered languages (1, 4, 9). Models can provide useful information about the potential success of intervention strategies; for example, they can inform about the total intervention strength needed to achieve the desired goal of language planners (10).

Modeling frameworks describing language shift can be roughly divided into two categories: agentbased and equation-based approaches. By simulating specific rules about individual behaviors, agent-based models often focus on elaborating the necessary interaction and transmission mechanisms (e.g., refs. 11–13), whereas by transforming these behaviors into mathematical principles, equation-based models usually concentrate on how to describe the shift dynamics at the population level (14). Following an influential paper by Abrams and Strogatz (15), most of the recent approaches have been equation-based. These studies have emphasized the importance of bilingual or diglossic equilibria (e.g., refs. 10, 12, and 16), as well as of the demography of speaker populations (e.g., refs. 17–19) for the shift dynamic. The behavior of these models is in part controlled by a "status" variable, summarizing the socioeconomic opportunities a language affords to its speakers. Researchers have estimated the strength of this parameter by model-fitting and successfully replicated census data recording the change in the numbers of speakers over time. Recently, Zhang and Gong (20) provided a parameterization of the status variable based on the concepts of language diffusion and inheritance.

Prochazka and Vogl's model (7) offers a different approach by providing a microscopic description of language shift on a fine spatial scale. The units of their model are 1-km2 grid cells characterized by the numbers of speakers of languages A and B they contain. Monolingual and bilingual speakers of the minority language B are aggregated into a single count. Changes in prevalence of each language in a cell are modeled based on the distribution of speakers in that cell and in its neighborhood (Fig. 1). In detail, to determine the number of speakers of minority language B in time step t, the total population size at this time (i.e., the combined numbers of speakers of both languages) in each cell is multiplied by the probability of speaking language B (see equation [3] in ref. 7). For a particular cell, this probability is given by the number of speakers of language B in the previous time step $t - 1$ and the influence of the speakers of language B in the neighborhood of this cell. The size of this neighborhood is parameterized by the parameter D, a constant that is estimated by fitting. Crucially, the entire model dynamic is governed by this space-dependent probability.

Prochazka and Vogl (7) apply their model to the case of language shift from Slovenian to German in southern Carinthia, for which spatially fine-grained census data exist for the periods 1880–1910 and 1971–2001. The fit between simulated and observed data, especially for the period 1880–1910, indicates that the prevalence of speakers of the same and of the competing language in a neighborhood are strong predictors of subsequent rates of language shift. The authors extend their framework and add what they call "habitat" conditions. If a cell is situated in an urban environment, or possesses a bilingual school or church in which the minority language is used, then this cell can have an increased probability of speaking the minority language. Interestingly, adding such habitat conditions did not improve their model fit for the period between 1880 and 1910; however, distinguishing between urban and rural areas provided a better explanation of the data in the period 1971–2001.

The result that bilingual schools have a negligible influence is perhaps less surprising in the period 1880–1910, because these schools tended to promote German (21). A hundred years later, the world has changed and enrolment rates in bilingual elementary schools are reported to be rising markedly, apparently reflecting Carinthian Slovene-identifying and other Slovenian-valorizing parents' wishes for their children to grow up bilingual (22). But how well can such factors be captured in a census measure of language shift, if the early censuses did not capture data on bilingualism, and if monolingual and bilingual Slovene speakers are aggregated into a single category for modeling purposes? As Prochazka and Vogl (7) recognize, state census data have their limitations. Unlike laboratory measurement, these data are selfreported classifications. Individual census returns claiming to speak (or not to speak) aminority language in a multilingual nation-state are liable to bias, both by self-perceived social identity and by expectations of the social and political purpose of the census itself. Thus, in the case of Slovenian in Carinthia, it has been estimated that the 1991 census underreported speaking knowledge of Slovenian by perhaps 50% (8). Despite these caveats, census data are often the only source of temporal and spatial information about language use at the population level. The strength of modeling approaches like Prochazka and Vogl's (7) is that they make predictions that can be tested with finer-grained sociolinguistic data, and which potentially lead to future lines of investigations. Additionally, thosemodels can be used to test whether the data at hand constitute a clear enough diagnostic indicator of the factors analyzed and therefore to inform future data collection.

In the same line, Prochazka and Vogl's (7) finding that in southern Carinthia, during 1971–2001, language shift was slower in urban than in rural areas is contrary to the expectation that the density and diversity of urban populations should accelerate shift (because of higher rates of mixing and the need for a lingua franca) (6). Historical rural–urban migration rates of Slovenian speakers might shed some light on this; but a recent sociolinguistic study suggests that social identity will be key to maintenance of Slovenian in predominantly German-speaking towns and cities. Where multilingualism is a sign of cosmopolitanism, young people brought up bilingual more willingly maintain and display their breadth of language skills (23). Prochazka and Vogl's (7) results emphasize the importance of a motivated community of speakers if an endangered language is to survive: use it, or lose it.

Understanding language shift is an interdisciplinary challenge, and modeling frameworks provide a tool for interpreting linguistic census data at the population scale. The model presented by Prochazka and Vogl (7) has an appealing simplicity and allows for the inclusion of various social factors, such as bilingual schools or urban environments. In addition to the generation of testable predictions, further developments may provide a platform to address the questions raised above, and also to engage with linguists' expectations for an effective dialogue as the modeling literature develops (24).

1 Grenoble LA, Whaley LJ (2006) Saving Languages: An Introduction to Language Revitalisation (Cambridge Univ Press, Cambridge, UK).
2 Krauss M (1992) The world's languages in crisis. Language 68:4–10.
3 Sutherland WJ (2003) Parallel extinction risk and global distribution of languages and species. Nature 423:276–279.
4 Amano T, et al. (2014) Global distribution and drivers of language extinction risk. Proc Biol Sci 281:20141574.
5 McMahon A (1994) Understanding Language Change (Cambridge Univ Press, Cambridge, UK).
6 Grenoble LA (2011) Language ecology and endangerment. The Cambridge Handbook of Endangered Languages, eds Austin PK, Sallabank J (Cambridge Univ Press, Cambridge, UK), pp 27–44.
7 Prochazka K, Vogl G (2017) Quantifying the driving factors for language shift in a bilingual region. Proc Natl Acad Sci USA, 10.1073/pnas.1617252114.
8 Priestly T (2003) Maintenance of Slovene in Carinthia (Austria): Grounds for guarded optimism? Can Slavon Pap 45:95–117.
9 Tsunoda T (2005) Language Endangerment and Language Revitalization (Mouton De Gruyter,

Berlin).
10 Kandler A, Unger R, Steele J (2010) Language shift, bilingualism and the future of Britain's Celtic languages. Philos Trans R Soc Lond B Biol Sci 365:3855–3864.
11 Castello, X, Eguıluz VM, San Miguel M (2006) Ordering dynamics with two non-excluding options: Bilingualism in language competition. New J Phys 8:308.
12 Minett JW, Wang WS-Y (2008) Modelling endangered languages: The effects of bilingualism and social structure. Lingua 118:1945.
13 Schulze C, Stauffer D, Wichmann S (2008) Birth, survival, and death of languages by Monte Carlo simulation. Commun Comput Phys 3:271–294.
14 Gong T, Shuai L, Zhang M (2014) Modelling language evolution: Examples and predictions. Phys Life Rev 11:280–302.
15 Abrams DM, Strogatz SH (2003) Linguistics: Modelling the dynamics of language death. Nature 424:900.
16 Mira J, Paredes ´A (2005) Interlinguistic similarity and language death dynamics. Europhys Lett 69:1031–1034.
17 Patriarca M, Heinsalu E (2009) Influence of geography on language competition. Physica A 388:174–186.
18 Kandler A (2009) Demography and language competition. Hum Biol 81:181–210.
19 Isern N, Fort J (2014) Language extinction and linguistic fronts. J R Soc Interface 11:20140028.
20 Zhang M, Gong T (2013) Principles of parametric estimation in modeling language competition. Proc Natl Acad Sci USA 110:9698–9703.
21 Fischer R (1986) The bilingual school of the Slovenes in Austria. J Multilingual Multicult Dev 7:187–197.
22 Busch B (2010) Slovene in Carinthian—Language beyond ethnic categories. Border dissolutions: Multilingualism, Transculturality and Education, eds Wintersteiner G, Gombos G, Gronold D (Wieser, Klagenfurt, Germany), pp 144–157.
23 Weichselbraun A (2014) 'People here speak five languages!': The reindexicalization of minority language practice among Carinthian Slovenes in Vienna, Austria. Lang Soc 43:421–444.
24 Wichmann S (2014) The challenges of language dynamics: Comment on "Modelling language evolution: Examples and predictions" by Gong, Shuai, and Zhang. Phys Life Rev 11:303–304.

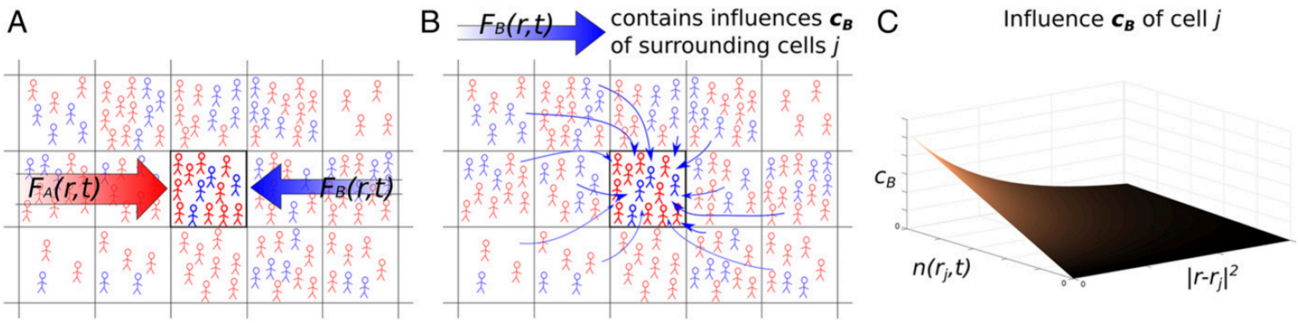

Fig. 1. (A) The probability of speaking language B in the focal cell (bold borders) is given by the number of speakers of language B in this cell (indicated by the color blue) and the influence of the speakers of language A and B, denoted by FA(r,t) and FB(r,t), respectively, in the neighborhood of this cell. (B) The strength of the influence FB(r,t) is determined additively by the influences cB(r,rj,n(rj,t),t) of the surrounding cells j. (C) The strength of the individual influences is depends on the distance jr-rjj between the focal cell and cell j and the number of speakers of language B in cell j

in a way shown.