Learnability pressures on language across multiple timescales

Molly Lewis
mll@stanford.edu
Department of Psychology
Stanford University

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

Abstract

Keywords: Linguistic Niche Hypothesis; language evolution

Introduction

Many facts about the social world are explicable only by considering the broader context: Why do gloves have five fingers? Why do tropical countries use lots of spices? The answers to these questions rely on taking into account properties of individual humans (hands have five fingers), and properties of the broader environmental context (tropical countries are hot, and spices kill bacteria; Sherman & Billing, 1999). A growing body of work has begun to also explain *language structure* by taking into account these same contextual factors. This work argues that language systems are the product of cognitive constraints internal to the language learner (Chater & Christiansen, 2010), as well as properties of the environmental context (Nettle, 2012) .

In this paper, we explore this second proposal—that properties of the environmental context shape language systems. This hypothesis, what has been termed the *Linguistic Niche Hypothesis* (Lupyan & Dale, 2010; Wray & Grace, 2007), suggests that language systems adapt to the pressures in the environment where a language is spoken. These pressures may come from a range of sources, including the climate, the length of the growing season, or the demographic characteristics of the people who learn and speak the language.

An important aspect of this hypothesis is the scale over which these pressures operate (McMurray, Horst, & Samuelson, 2012; Blythe, 2015). There are at least two scales at which language can be studied: individual utterances, which change over the course of moments in a communicative interaction, and language systems, which change over the timescale of years. An established body of work supports the claim thats speakers adapt to the context at the scale of individual utterances, both in terms of the properties of the listener and the physical environment (e.g., Schober, 1993; Frank & Goodman, 2012). The present hypothesis explores the more controversial claim that environmental pressures shape language at the scale of language systems.

There are a number of pieces of evidence that environmental factors may indeed shape language systems. At the lowest level of the linguistic hierarchy, languages with larger populations are claimed to have larger phonemic inventories (Atkinson, 2011; Hay & Bauer, 2007), but shorter words (Wichmann, Rama, & Holman, 2011). Speakers with more second language learners have also been suggested to have a lower type-token ratio of lexical items (Bentz, Verkerk, Kiela, Hill, & Buttery, 2015). At the level of morphology, evidence suggests that speakers with larger populations tend to have simpler morphology (Lupyan & Dale, 2010; Bentz & Winter, 2013). Finally, there is also evidence that population size may influence the mappings between form and meaning. In particular, this work suggests that languages tend to map longer words to more complex meanings (Lewis, Sugarman, & Frank, 2014), but that this bias is smaller for languages with larger populations (Lewis & Frank, 2016).

The plausibility of the Linguistic Niche Hypothesis depends largely on the presence of a possible mechanism linking environmental features to aspects of language systems. A range of proposals have been suggested (Nettle, 2012). For example, one possibility is that children (L1) and adult (L2) language-learners differ in their learning constraints. In particular, children may be better at acquiring complex morphology than adults, and so languages with mostly children learners may tend to have more complex morphology. A second possibility is that speakers in with less dense social networks have less variable linguistic input, and this leads the language system to have more complex morphology.

Testing these mechanisms is empirically challenging, however. Because there are many factors that shape a linguistic system, large datasets are needed to detect an effect of environmental factors. In addition, many languages are non-independent because of genetic relationships and language contact, and so data from a wide range of languages are needed to control for these moderators. Third, the scale of this hypothesis makes it difficult to directly intervene on the mechanism. Finally, the hypothesized mechanisms are somewhat underspecified, and the dynamics of these different factors may be complex, trading-off with each other in non-obvious ways (e.g. Wichmann et al., 2011).

In this work, we try to address these challenges by clarifying the empirical landscape. We do this by aggregating across datasets that find covariation between environmental variables and linguistic structure. This serves two purposes. First, it allows us to examine the relationship between the same set of environmental predictors across a range of linguistic features. And, second, it allows for the same analytical techniques and areal controls to be used across datasets. By addressing these inconsistencies, we are in a better position to more directly compare relationships between environmental and linguistic features. Importantly, a more coherent picture of the empirical landscape may provide insight into the mechanism linking language systems to their environments.

We also more directly address the question of mechanism by examining variability in the mean age of acquisition of

words for L1 learners across languages. Evidence that this variability is related to an aspect of the linguistic system (such as number of phonemes) would suggest that L1 learners, and not L2 learners, are the relevant environmental factor shaping that aspect of the linguistic system.

In what follows, we first present a set of analyses examining the relationship between environmental and linguistic features using the same analytical techniques (Analysis 1). In Analysis 2, we examine the relationship between a languages mean age of acquistion and aspects of the linguistic system.

Analysis 1: Environmental pressures on language systems

(Jaeger & Tily, 2011) (Moran, McCloy, & Wright, 2012)

Analysis 2: L1 learning and language systems

(Łuniewska et al., 2015)

Other work has empirical + modeling Some attempts (Silvey, Kirby, & Smith, 2015) (Perfors & Navarro, 2011)

(Wichmann et al., 2011) (?, ?) (Smith & Wonnacott, 2010) (Slobin & Bever, 1982)

(Sapir, 1912) (REALI, CHATER, & CHRISTIANSEN, 2014)

(Lupyan & Dale, n.d.)(Lupyan & Dale, 2010) (Kirby, Cornish, & Smith, 2008) Meaning. (Silvey et al., 2015) (Perfors & Navarro, 2011) Critically,

References

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from africa. *Science*, *332*, 346–349.

Bentz, C., Verkerk, A., Kiela, D., Hill, F., & Buttery, P. (2015). Adaptive communication: Languages with more non-native speakers have fewer word forms.

Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, *3*, 1–27.

Blythe, R. A. (2015). Hierarchy of scales in language dynamics. *arXiv preprint arXiv:1505.00122*.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*, 1131–1157.

Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998–998.

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

Jaeger, T. F., & Tily, H. (2011). On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 323–335.

Figure 1: Object stimuli used in the Experiment. The objects are sorted from least complex (top left) to most complex (bottom right) based on the complexity norms in Lewis et al. (2014). Each row corresponds to a quintile.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10686.

Lewis, M., & Frank, M. C. (2016). Learnability pressures influence the encoding of information density in the lexicon learm. In *The evolution of language conference.*

Lewis, M., Sugarman, E., & Frank, M. C. (2014). The structure of the lexicon reflects principles of communication. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society.*

Łuniewska, M., Haman, E., Armon-Lotem, S., Etenkowsk, B., Southwood, F., Pomiechowska, A., ... others (2015). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods*.

Lupyan, G., & Dale, R. (n.d.). The role of adaptation in understanding linguistic diversity. *The Shaping of Language: The Relationship between the Structures of Languages and their Social, Cultural, Historical, and Natural Environments*.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*, e8559.

McMurray, B., Horst, J., & Samuelson, L. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological review*, *119*, 831.

Moran, S., McCloy, D., & Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language*, *88*, 877–893.

Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1829–1836.

Perfors, A., & Navarro, D. (2011). Language evolution is shaped by the structure of the world: An iterated learning analysis. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 477–482).

REALI, F., CHATER, N., & CHRISTIANSEN, M. H. (2014). The paradox of linguistic complexity and community size. In *The evolution of language: Proceedings of the 10th international conference* (pp. 270–277).

Sapir, E. (1912). Language and environment1. *American Anthropologist*, *14*, 226–242.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1–24.

Sherman, P. W., & Billing, J. (1999). Darwinian gastronomy: Why we use spices. *Bioscience*, *49*, 6.

Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive science*, *39*, 212–226.

Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, *12*, 229–265.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*,

116, 444–449.

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

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*, 543–578.