Proceedings of the Society for Computation in Linguistics

Volume 1 Article 37

2018

A Bayesian Investigation of Factors Shaping the Network Structure of Inflection Class Systems

Jeff Parker
Brigham Young University, jeff_parker@byu.edu

Robert Reynolds
Brigham Young University, robert_reynolds@byu.edu

Andrea D. Sims
The Ohio State University, sims.120@osu.edu

A Bayesian investigation of factors shaping the network structure of inflection class systems

Jeff Parker°, Robert Reynolds°, Andrea D. Sims*
°Brigham Young University *The Ohio State University

In this paper we use a Bayesian, agent-based model to explore the emergence of allomorph distributions in inflection class (IC) systems. It has long been understood that irregular inflection occurs mainly among high token frequency lexemes because high frequency leads to word-specific learning, allowing certain lexemes to resist analogical pressure. Over time, these lexemes become ‘stranded’ in low type frequency classes as less frequent lexemes shift class membership. Crucially, these classes partly overlap but do not collapse with high type frequency classes, and as a result detract from speakers’ ability to predict a word’s inflection class membership. Stump and Finkel (2013) extract an empirical generalization that they call Marginal Detraction (MD): low type frequency classes contribute more to the complexity of an IC system than high type frequency classes do. This is represented as negative slopes in Figure 1 based on data from Sims and Parker (2016). Here, IC system complexity (operationalized as conditional entropy) is defined as the average uncertainty associated with one inflected form of a lexeme, given knowledge of one other form of the same lexeme.

What remains unclear is why some languages exhibit Marginal Detraction and others do not. After all, if MD is an emergent distributional property that results from individual words resisting analogy based on their high token frequency, we might expect to find it in every IC system to the extent that word frequency distributions are universally Zipfian. Since we seem not to, the resistance of individual lexemes to analogy seems not to be the full story. In this paper we computationally test a hypothesis from Sims and Parker (2016) that there is a further, systemic pressure at work. Specifically, we explore the idea that differences in the structure of overlaps among allomorphs in IC systems, which creates analogical pressure, may play a role in the historical process that leads to MD is some, but not all, languages with IC systems. We thus approach the MD pattern from a typological angle, asking what kinds of inflectional systems might lead to its emergence.

We created a multi-generational agent-based model in which each agent was equipped with a Bayesian learner. In the model, a hypothesis *h* is a probability distribution over the set of allomorphs for a given morphosyntactic property set (MSPS), conditioned on the known allomorph of some other MSPS. In a pseudo-Russian language a hypothesis might be as in (1). A prior probability is assigned to each hypothesis in the hypothesis space based on observations of other lexemes meeting the conditioning environment (‘neighbors’), e.g., lexemes having ACC.SG = /u/. The prior thus reflects analogical pressure in the model. A production *d* consists of a lexeme and a pair of allomorphs, representing the observation and conditioning environment. In pseudo-Russian, a production might look like (2).

(1) h_1 = NOM.SG | (ACC.SG = /u/) : {/NULL/$_{p = 0.3}$, /a/$_{p = 0.4}$, /o/$_{p=0.3}$}

(2) d = {KNIGA, [NOM.SG = /a/ | ACC.SG = /u/]}

In Bayesian inference, the posterior probability of hypotheses reflects a balance between observed productions and expectations (here, neighbor behavior as encoded in prior probability). However, crucially, in our model we held the number of productions constant for all lexemes, thus removing the effects of token frequency. This allowed us to directly investigate how the overlap of allomorphs across classes affects analogy within the system.

Input to the model: We created ten artificial IC systems that each contained six MSPS and 24 classes. Treating IC systems as undirected graphs in which classes are nodes and edges connect nodes that share allomorphs, we manipulated input node degree (low 1, high 23) and mean edge weight (low 1.55, high 4.13). Node degree is the number of classes with which a target class shares one or more allomorphs. Edge weight is the number of allomorphs shared between two connected classes. Within each input system, every class had equal degree and mean edge weight so that removing any one class would have the same effect on system complexity as removing any other class. Node degree and edge weight represent two dimensions of analogical pressure within an IC system.

Proceedings of the Society for Computation in Linguistics (SCiL) 2018, pages 223-224.
Salt Lake City, Utah, January 4-7, 2018

Model parameters: 1,000 lexemes were assigned to ICs based on a distribution designed to simulate a natural language (based on Russian nouns). In each generation, 50 adult agents produced forms as in (2) and 50 child agents listened to 100,000 productions from each of three randomly selected adult agents. At the end of the listening stage, child agents applied the Bayesian learning algorithm, predicting inflected forms for every MSPS of every lexeme, with a winner-takes-all principle. Child agents then matured into adult agents for the next generation and produced forms based on sampling from the output of their Bayesian inference process. We ran the model for 10 generations, with 6 runs per input condition.

Results showed that the way the IC systems were restructured across generations depended on input degree and edge weight. In systems with high input degree and high input edge weight, the class systems collapsed within the few generations, but in systems with low degree or low edge weight, the systems restructured. There was minimal variation across runs with the same input. Most importantly among systems that restructured, input degree predicted whether the Marginal Detraction pattern emerged. Input systems with low degree produced Marginal Detraction in the output to the model (shown as negative slope in Figure 2). In other words, low type frequency ('irregular') classes that contribute disproportionately to the complexity of an IC system were more likely to emerge when classes were more distinct overall.

Our results suggest that typological differences in the internal structuring of IC class systems may affect the likelihood that a system will exhibit Marginal Detraction. This suggests that not only lexeme token frequency, but the internal structure of IC systems plays a role in the typological patterns that arise among IC systems. The way in which classes are embedded in an IC system – specifically, the pattern of overlaps among classes – is a potentially important and independent determinant of how IC systems evolve, including the relationship between regulars and irregulars.

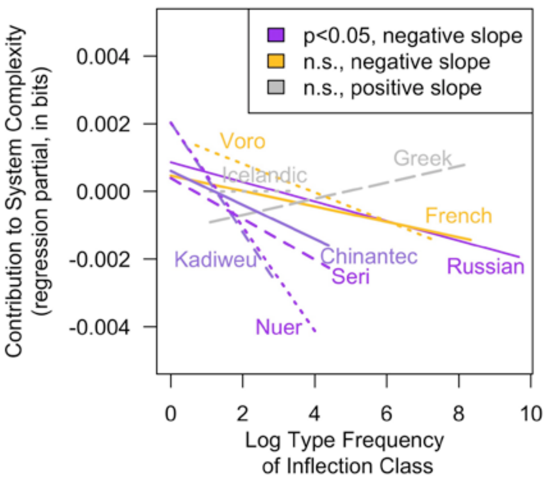

Figure 1: Regression slope for type frequency as a predictor of system complexity in nine natural languages

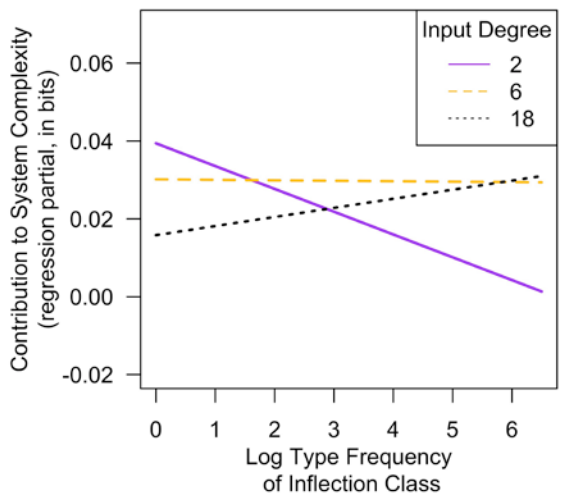

Figure 2: Model output. Regression slope for type frequency as a predictor of system complexity (non-collapsing systems only), grouped according to input node degree

References

Ackerman, Farrell & Robert Malouf. 2013. Morphological Organization: the Low Entropy Conjecture. *Language* 89(3). 429–464.

Brown, Dunstan & Andrew Hippisley. 2012. *Network Morphology*. New York: Cambridge University Press.

Sims, Andrea D. & Jeff Parker. 2016. How inflection class systems work: On the informativity of implicative structure. *Word Structure* 9(2). 215–239.

Stump, Gregory T. & Raphael A. Finkel. 2013. *Morphological Typology: From Word to Paradigm*. New York: Cambridge University Press.