Emergent idiosyncracy in English comparatives

Brian Smith and Claire Moore-Cantwell
UCSC, UConn

Introduction: Although speakers have knowledge of phonological trends across the words of their lexicons (Ernestus and Baayen 2003, Hayes et al 2009, et seq.), in many cases individual items behave idiosyncratically (Zuraw 2016). We present the results of a computational model which uses UR constraints (Pater et al. 2012) to represent lexical idiosyncracy. We demonstrate that across generations, the behavior of individual lexical items can diverge. We specifically examine the case of the English comparative, which can be realized with *-er* (happier) or *more* (more happy). Although the choice between *-er* or *more* is influenced by phonological factors, it is ultimately idiosyncratic to particular adjectives. Crucially, our model requires storage of *more* forms in addition to storage of *-er* forms. This expansive view of what's stored in the lexicon builds on work in different phonological domains, such as French liaison (Bybee 2000), and English binomials (Mollin 2012), both of which argue for the storage of multi-word sequences.

Lexical idiosyncrasy in the English comparative: The choice between *more* and *-er* is governed by both phonological and syntactic factors (Hilpert, 2008). Monosyllabic words prefer *-er* (taller > more tall), while 3+ syllable words prefer *more* (more unimpeachable > unimpeachabler). The adjective's final segment (Kytö & Romaine 1997, Lindquist 2000), and stress (Leech & Culpeper 1997) are also predictive.

Adj.	P(*-er*)	Total (*more* + *-er*)
lonely	0.75	163
deadly	0.42	355
timely	0.11	167
worldly	0.05	95
likely	<0.01	21,899

However, even when these phonological factors are held constant, adjectives vary greatly in their rate of appearance with *more* vs. *-er*. The table shows five adjectives, all initially stressed and ending in [li], whose rates of taking *-er* vary from < 1% to 75%. Counts come from COCA (Davies 2008–).

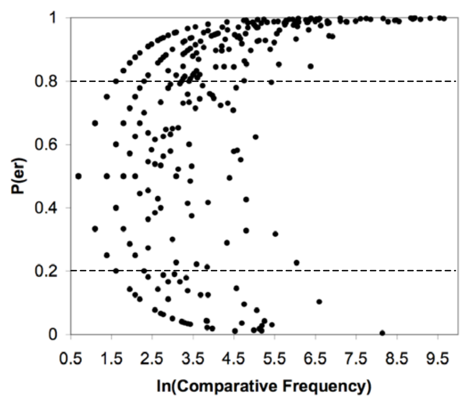
Historical development: Kytö and Romaine (2000) show that, over time, individual adjectives have become more strongly idiosyncratic. In late Middle English, most adjectives used *-er* about 30% of the time. Since then, *-ous* final adjectives have drifted towards *more* (now at 0% *-er*) and *-li/-y* final adjectives have drifted towards *-er* (now at about 80% *-er*). Although *-ly* final adjectives now generally prefer *-er*, exceptions remain, such as *likely* above. The goal of our model is to capture this arbitrary drift towards either *-er* or *more*.

Frequency effects: The results of this drift can be seen in the synchronic grammar, in which high-frequency adjectives prefer *more* or *-er* nearly categorically. For low frequency adjectives, the rate of *-er* is closer to 50%. This is shown in the left-hand figure below, from Boyd (2007), which plots comparatives from the British National Corpus. We argue that this effect of frequency is a result of speakers' ability to explicitly store the comparative form(s) of an adjective. As a speaker is exposed to more instances of a given comparative, that comparative will be stored more strongly in the speaker's lexicon. The speaker will be less likely to forget it, and more likely to use it rather than compose the comparative anew.

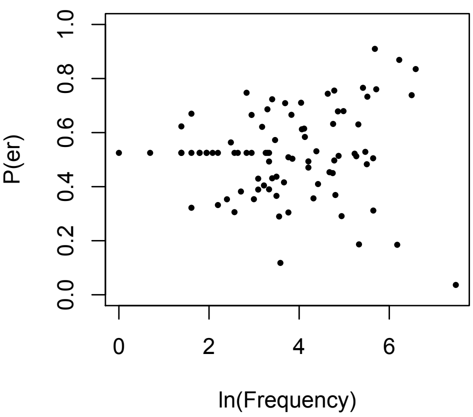
Model description: Storage of comparatives in the lexicon is modeled with UR constraints of the form: PRETTY + COMPARATIVE → /pɹɪɾiɹ/ (Boersma 2000, Pater et al. 2012, Smith 2015). This constraint assigns a violation whenever the word *pretty* in the comparative is realized with *more* rather than with *-er*. UR constraints are situated within a MaxEnt grammar (Goldwater & Johnson 2003), which also contains two phonological constraints – one preferring *more* and one preferring *-er* – which stand in for the complex array of phonological variables discussed above. UR constraints are induced, and sometimes forgotten, during learning. Learning proceeds gradually, sampling one word at a time based on lexical frequency. If the learner makes an error on a word, it uses the delta rule (Rumelhart & McClelland 1986) to update existing constraints, and also creates a new UR constraint if necessary. Weights of UR constraints decay over the course of learning, so that ones that are never updated eventually dissappear. UR constraints affiliated with lower frequency lexical items are more fragile, since opportunities to update on those lexical items are rare.

The starting data was a toy dataset of 100 lexical items, varying in frequency according to a zipfian distribution. The first generation saw data in which every word had a 50% chance of appearing with *-er*. Each generation trained for 10,000 updates, with a learning rate of 0.1 and a decay rate for UR constraints of 0.0001 (the weight of a UR constraint that is not updated for 5 updates would decay by 0.0005). UR constraints, when induced, had a starting weight of 1. Each generation trained on the end-state of the previous generation's learning, and the model ran for 20 generations. The right-hand graph below shows the results of the final generation.

BNC (Boyd 2009)



Model



Since UR constraints of high-frequency lexical items are updated often, small mismatches between the input and the result of learning at each generation magnify into very pronounced differences between individual high-frequency lexical items. Lower-frequency lexical items remain undecided, since their UR constraints tend to decay away.

Conclusion: We model two aspects of lexical idiosyncracy in English comparatives: its historical development, and its interaction with frequency. Crucial to the success of our model is the use of UR constraints to 'store' both *-er* and *more* forms of the comparative. Because of this storage ability, lexical idiosyncracy emerges across multiple generations of learning.