Self-organization of the suffix system: productivity as competition

In recent years, it has been shown that certain subsystems of human languages can be profitably investigated in terms of self-organizing or emergent systems (Baayen (1993), Anshen and Aronoff (1999), Albright (2002), Plag and Baayen (2009)). This investigation examines affix productivity as a system of competition among morphemes. Three fundamental aspects of language force affixes to compete in a "survival of the fittest": the self-perpetuation of suffixes through productive derivation, a general intolerance for synonymy, and the continual introduction of random elements into the system. Competition arises when two affixes depend on productivity within the same domain; this competition is ultimately resolved either through the loss of productivity of one competitor or the development of subdomains (phonological, morphological, etc.) where different suffixes can dominate. Productivity will be measured in a novel way: using statistical analysis that incorporates Estimated Total Matches from Google search results that are gathered in an automated fashion. This method differs from those that use *hapax*-based measures (e.g. Baayen 1993), though they are trying to capture the same intuitive point: that some affixes are more active than others in forming new words.

The suffix pair *-ic* and *-ical* is composed of two suffixes that are synonymous, and therefore in direct competition. Using Webster's 2nd dictionary, a total of 11966 stems of English words ending in either *-ic* or *–ical* were identified. Next, queries were executed (using the Google Search API) for words ending in both *-ic* and *-ical* with each stem. For some stems, a Google search will record a high number of hits for both *-ic* and *-ical* words (e.g. *historic* and *historical*), while for others, one suffix will yield far more hits. The word with the most hits in each pair was considered the 'winner' for that pair. Ninety percent of all comparisons yielded a winner by a margin of at least one order of magnitude. Overall, this comparison produced 10613 *-ic* winners vs. 1353 *-ical* winners with an overall ratio of 7.84 in favor of *-ic*, demonstrating conclusively that—by this measure of productivity—*-ic* is more productive than *-ical* overall.

However, within a certain domain, *-ical* is potentiating (Williams 1981). To investigate this, stems were sorted into left-to-right alphabetical neighborhoods of one to five letters, e.g. all stems ending in *-t-* (4166 members), or all stems ending in *-graph-* (294 members). We found that the only subset of words favoring *-ical* with at least a dozen members is *-olog-*; for this subset only, *-ical* is the winner over *-ic* (e.g. *psychological* over *psychologic*) by a ratio of 8.30, almost the exact reverse of the ratio of the full set (7.84 in favor of *-ic*). In other words, using this technique, we find that, although *-ic* is more productive than *-ical* overall, the reverse is true for words ending in *-olog(ical)*.

Although *-olog(ical)* forms a large subset (475 members), no other large sets in the system have resisted the overall trend favoring *-ic*. However, the *-olog(ical)* subset is unique in another way: the string 'olog' is strikingly distinct. For example, there are 79 stems ending in *-rist-*, but 660 stems ending in the substring *-ist-*, and this number jumps to 4166 stem ending in *-t-*. Thus, *-rist-* makes up only 1.9% of stems ending in *-t-*, leaving many similar neighbors. On average, a neighborhood of length 2 (e.g. *-st-*) only accounts for 27.84% of the words in the corresponding neighborhood of length 1 (e.g. *-t-*), while a neighborhood of length 4 makes up just 10.47% of its length-2 set. However, even at length 4, the *-olog-* subset still makes up 66.62% of its length 3 (*-g-*) set. This means that 66% of words ending in *-gic(al)* also end in *-ologic(al)*, which is the strongest of all sets by a wide margin (the next strongest is *-graph-* at 34%). Thus, *-olog(ical)* is a subsystem that is not only sufficiently large, but also is distinctly uniform, leaving it uniquely suited to sustain itself in spite of patterning inversely to all other *-ic*/*-ical* pairs.

The pair of *-ize* and *-ify* represents another rivalry. Both suffixes convert nouns and adjectives to verbs, with the meaning "render, make, convert into" (Marchand 1969). Like, *-ic* and *-ical*, there is usually a strong preference for one or the other for a given stem. Out of 2636 unique stems ending in either *-ize* or *-ify*, 2217 favored *-ize* in head-to-head competition, while 419 favored *-ify*, yielding an approximate 5:1 ratio in favor of *-ize*. However, these results are governed by phonological restrictions. If these results are reorganized according to the number of syllables in the stem, a different pattern emerges. We find that, while polysyllabic stems still favor *-ize* (2127 *-ize* vs. 89 *-ify*), monosyllabic stems overwhelmingly favor *-ify*. In this subset, there were 322 *-ify* winners versus only 68 *-ize* winners; thus, by a ratio of nearly 5:1, *-ify* is favored over *-ize* in this domain.

As with *-ic* and *-ical*, we have a large and clearly-defined subset of all possible words. While *-ic* and *-ical*'s subsets were defined along morphological boundaries, the *-ize* and *-ify* subsets are restricted phonologically. Each achieves the same result: two competing suffixes are able to coexist in a system that does not tolerate synonymous affixes by finding a niche within a subset of their domain.

Selected References

Albright, Adam. 2002. The Identification of Bases in Morphological Paradigms. Ph. D. thesis, UCLA.

Anshen, Frank & Mark Aronoff. 1999. Using dictionaries to study the mental lexicon. *Brain and Language* 68.10.

Baayen, R. Harald. 1993. On frequency, transparency, and productivity. In Booij, G. and J. van Marle (eds.) *Yearbook of morphology 1992*. Dordrecht: Kluwer. 181-208.

Bloomfield, Leonard. 1933. *Language*. Revised from 1914 edition. New York: Holt.

Booij, Geert. 1987. The Reflection of Linguistic Structure in Dutch Spelling. In Orthography and Phonology, Luelsdorff, Philip A. (ed.), 215 ff.

Booij, Geert. 2005. The Grammar of Words: An Introduction to Linguistic Morphology. Oxford: Oxford University Press.

Bart de Boer (1999) Evolution and self-organisation in vowel systems Evolution of Communication 3 (1), 79–103

Edelman, G. M. 1987. Neural Darwinism: The Theory of Neuronal Group Selection. New York: Basic Books.

Fogel, David B. 1993. Evolving Behaviors in the Iterated Prisoner's Dilemma. *Evolutionary Computation*, 1(1): 77-97.

Nelson, Richard R. & Sidney G. Winter. 2002. Evolutionary Theorizing in Economics. *Journal of Economic Perspectives*, 16(2): 23-46.

Goldberg, David E. 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley.

Marchand, Hans. 1969. The categories and types of present-day English word-formation. A synchronic-diachronic approach. Munich, Germany: Beck.

Plag, I. and Baayen, R. H. (2009). Suffix ordering and morphological processing. Language, 85, 106-149.

Plag, Ingo (1999). Structural constraints in English derivation. Berlin: Mouton de Gruyter.

Steels, L. (1997) The Synthetic Modeling of Language Origins. In: Gouzoules, H. (ed) Evolution of Communication, vol. 1, nr. 1, pp. 1-34, Amsterdam: John Benjamins Publishing Company.

Williams, Edwin. 1981. Argument structure and Morphology. *Linguistic Review*, 1:81-114.