

### How productive were English derivational affixes in the late 20<sup>th</sup> century?

Baayen and Renouf tried to determine the level of productivity of five English derivational affixes. At the time of their research, various proposals had already been put forth regarding productivity of the affix (for example, *-ly* was no longer considered productive by Cannon). Baayen and Renouf find flaws in the various approaches taken by other researchers, especially those that are dictionary based.

Using a corpus containing articles from the Times (newspaper of London) collected over 46 months (in total), they process some 80-million words and look at lexical innovation across two broad themes:

1. How productive (statistically) are the five affixes in question when it comes to creating new words? The researchers conclude that, contrary to popular belief, the use of *-ly* as a suffix to form new words is alive and well (at the *Times*, at least). They spend a goodly portion of their paper discussing the statistical formulae used and the rationale for choosing hapax legemena as the definition of “new word” (neologism) for their study.
2. Given the data from the corpus of 80-million words, specifically, the numbers of word types with each of the suffixes in question, what can be said about the structural properties of the base words? Are some types of base words more productive than others? The researchers find that Latinate affixes *-in* and *-un* are less productive than their Germanic counterparts

They also took the opportunity to dispute claims by the structuralists about the relationship between the number of constraints, the kind of base words and the degree of productivity.

\*\*\*

This research paper caught my eye at first because it involved a newspaper. Based on the first part of the title, I was expecting a lot more insight into the evolution of language over the course of time, at the *Times*. But of course the main focus of the paper was on statistical analysis over a rather short period, really, not even two full years, focused on lexical innovation with respect to use of affixes. It would have been interesting if the researchers had included some actual text examples, as full sentences, from the source material, along with the statistical information (they do give some examples later in the paper, but only of single words). And examples later in the article are from other researchers’ material.

On another level, I found the topic interesting because I’m trying to visualize how such a study must actually have occurred. Some of the implementation details of how the corpus was built would have been interesting, and that’s also what I thought I was going to be reading. After the mention of the AVIATOR project, the Times, how frequently articles were captured, and so on, I was hoping to read some of the implementation details. For example, it’s not clear how the “filters” mentioned early in the paper were actually used. How were words “registered” in the database? What does that mean, specifically?

Despite these initial missteps in my understanding of the nature of the paper, the topic of lexical innovation is interesting in its own right. Since languages evolve and change over time, an examination of lexical innovation with respect to affixes might be an indicator of that.

### **What conclusion does the article draw?**

From the highest level, Baayen and Renouf conclude that new words are created all the time, and that affixes long thought dead are highly productive. They point out that “the most productive affixes...[-ly, -ness, un-] ...give rise to lexical innovations that occur in increasing numbers over the months of sampling.”

Another conclusion is that base words *do* determine productivity to some extent, but that semantics is also a key driver. As the authors state, “the function of word formation is to convey (particular shades of) meaning, not simply to produce forms with a particular structure.”<sup>1</sup> For -ly, adverb base words proved most productive (1278/1362 hapax leg.), while for -ness, monomorphemes were most productive. The prefix un- was most productive with de-adjectival bases (646/659 hapax leg.), the use of un- seems to have supplanted the use of -in in most cases.

Finally, the researchers also tested their suppositions (by interpolating the number of types) using the Cobuild/CELEX database, which more or less confirmed their findings: -ness and -ity productivity curves “do not differ substantially,” while for un- and in- “curves diverge substantially.” They discovered, unexpectedly, that negative prefixation occurred much more in Cobuild than in the *Times*.

### **Do you find the authors’ approach satisfactory? If not, how else would you do it?**

From a purely statistical analysis perspective, Baayen and Renouf’s seems to make sense, although I was a bit surprised about the changes in the data collection after the first 15 months. Your explanation during class (the “back-story” about the AVIATOR project) helped clear that up, since it seemed really odd to modify the parameters of a study in mid-stream.

But as we discussed in class (and as Baayen and Renouf point out in their conclusion), the data presented does not take into account the role of the authors and editors at the *Times*. In addition, printing, typographical, and cost constraints, as well as technological innovation in that industry are likely all factors influencing the words chosen in any given moment. It would be interesting to try this study again, perhaps at a different publication, but use unedited versions of raw texts prior to editing and publication. Such a study might provide an interesting contrast to Baayen and Renouf’s study.

---

<sup>1</sup> “Chronicling the Times” p. 90, in discussion of the shortcomings of the ‘structuralist approach to morphological productivity’