

Mark Lindsay, mark.lindsay@stonybrook.edu
Stony Brook University

Fabio Montermini
Nabil Hathout
Université de Toulouse

Department of Linguistics
Stony Brook University
Stony Brook, NY 11794
USA

CLLE-ERSS
Université de Toulouse Le Mirail
Toulouse, France

Scientific Report

My dissertation work examines morphological productivity as a system of competition among morphemes, mirroring some aspects of evolutionary theory. Productive elements of a language survive or perish due, in part, to circumstances of environment and competition; the focus of the project will be on the competition among productive derivational affixes. Such competition arises when two affixes depend on productivity within the same domain (i.e. the same words in the same environments). These experiments will show that this competition is ultimately resolved either through the loss of productivity of a competitor or the development of subdomains where different affixes can dominate.

The primary aim of this short visit was to explore this rival suffix productivity in English and the Romance languages (especially Italian and French) and to develop my methodologies for measuring productivity quantitatively, particularly using freely available resources from the World Wide Web.

Description of work carried out

During my time in Toulouse, I met primarily with Fabio Montermini, Jesse Tseng, and Nabil Hathout. I stayed in an office at the CLLE-ERSS during the two weeks, and we discussed my research on suffix productivity on two fronts: first, we examined my methodology. I used Google search results and Estimated Total Matches (ETM) as components of my synchronic investigation of productivity of English and Romance suffixes. Nabil Hathout had also been working on mining information from the web to examine productivity. We found that we each were having similar difficulties on many fronts, but Nabil gave me some insight into the Google Web1 T5 N-Gram corpus data, and also pointed me towards a somewhat more traditional, freely available source: the uk/fre/de/itWaC corpora. I have already acquired these corpora and have parsed portions of the frWaC corpus for analysis, which I am carrying out with Mark Aronoff (Stony Brook) and Delphine Tribout (Paris).

Jesse Tseng and I examined my historical data for *-ize/-ify* and *-izzare/-ificare* in Latin, which came from the Oxford English Dictionary (OED) and the Oxford Latin Dictionary (OLD); we focused on the Latin data in particular and cross-referenced it with data with a web source called Perseus. We found that my initial Latin investigation agreed with the data from Perseus; this source also provided quite a bit more detail on these suffixes. We also began looking at two other related issues. First, the role of *en-* prefixed words in English in the

competition between *-ize* and *-ify*, and second, the degree (if any) of productivity of *-ize/-ify* equivalents in German. On the latter point, initial data indicate that German *-izieren* and *-ifizieren* are not productive and are only found in borrowed words.

I presented two talks, on 24 May and 1 June. The first was given for graduate students, along with Fabio and Nabil, about Lindsay and Aronoff (2012)'s "Natural selection in self-organizing morphological systems". I received a great deal of feedback and some possible areas of further investigation into the rival suffixes *-ic* and *-ical*, including possible orthographic factors and alternate cases. The second talk was a departmental seminar, titled "The emergence and evolution of suffix rivalry", about the evolution of rivals *-ity* and *-ity*, and *-ize* and *-ify*. This talk led to a discussion with Daniel Huber about his work on English *-ize*, which had not yet been published. He has analyzed the relationship between *-ize* and stress, as well as other semantic factors. We may collaborate in the future on this work.

Description of main results

Italian and English Historical Data on *-ize* and *-ify*

Lindsay (2012) investigated suffixes *-ize* and *-ify* in English, as well as their equivalents in French, Spanish, and Portuguese (*-iser/-ifier*, *-izar/-ificar*, and *-izar/-ificar*, respectively). Synchronic evidence was shown that these suffixes are distributed similarly across the lexicons of these languages with respect to prosodic structure. In particular, *-ify* forms strongly favor words with monosyllabic stems, while *-ize* forms strongly favor words with polysyllabic stems.

We see a striking similarity in the relative distributions of *-ize* and *-ify* forms across all four languages; Figure 1 below shows the percent of all words in each of the four languages where *-ize* is preferred, organized by number of stem syllables:

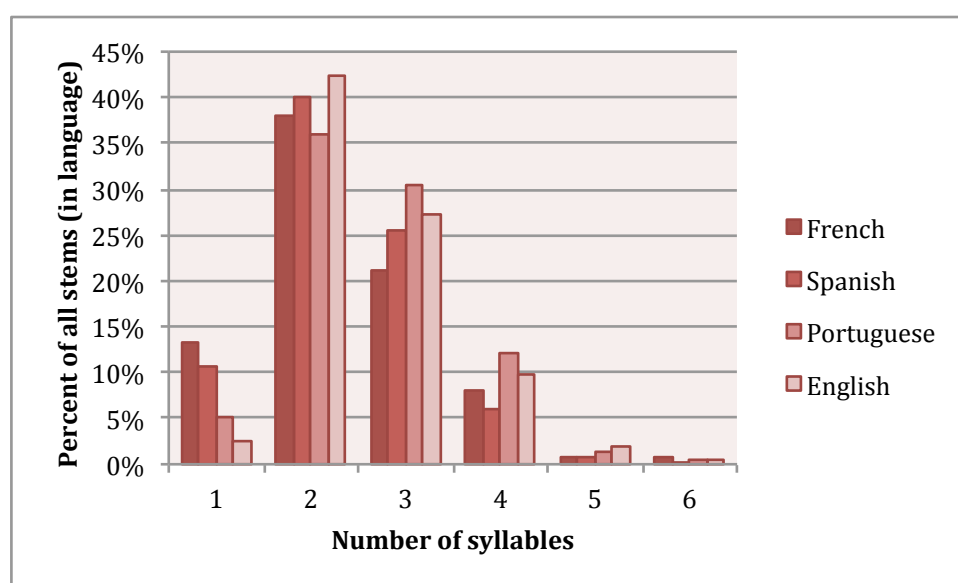
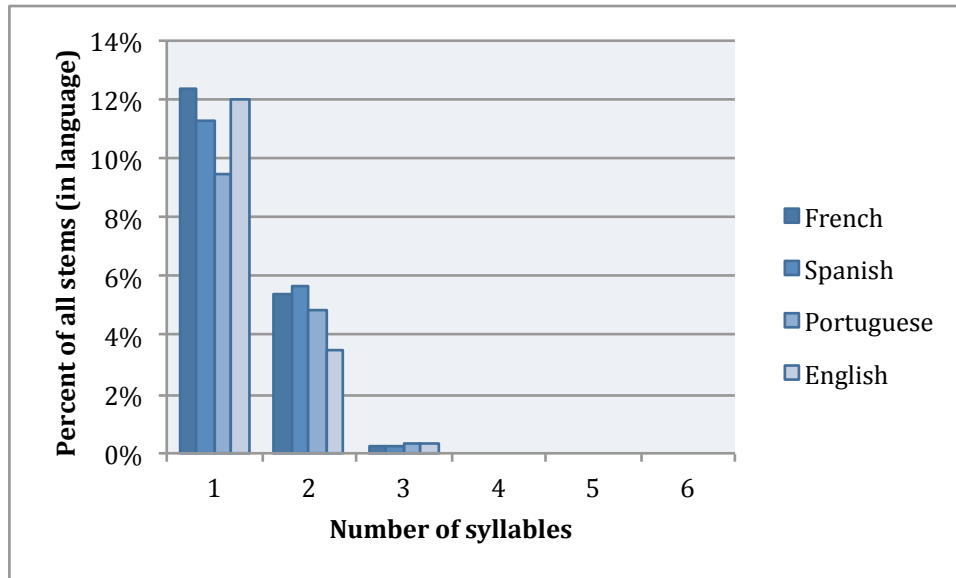


Figure 1: Relative distributions of *-ize* across languages.

Here we see a peak at two syllables and a significant number of words in the 3-6 syllable range. This contrasts with Figure 2 below, which shows the percentage of *-ify* words:

**Figure 2: Relative distributions of *-ify* across languages.**

Here we can clearly see the preference for monosyllabic stems, which drops off rapidly as the number of syllables increases.

In Toulouse we looked closely at the same phenomenon in Italian and compared it to English, both synchronically and diachronically, using data from the Grande Dizionario Italiano Dell'Uso (2003), available electronically.

In English, the emergence of *-ize* and *-ify* as productive patterns was measured using data (particularly first citation) in the Oxford English Dictionary (OED). As we see in Figure 3, whole-word borrowings (from French and Latin) preceded English derived forms; both *-ize* and *-ify* derivations became productive around the 16th Century.

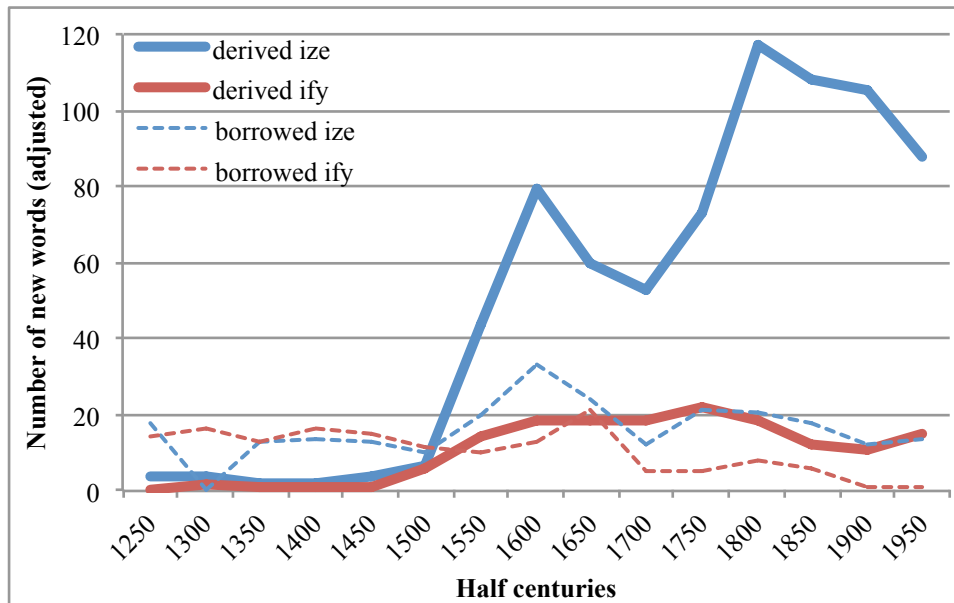


Figure 3: emergence of productivity of *-ize* and *-ify* in English.

Although both the English suffixes and Italian *-izzare* and *-ificare* trace their roots to Latin, it appears from our initial data that it was not until later that the Italian suffixes became productive.

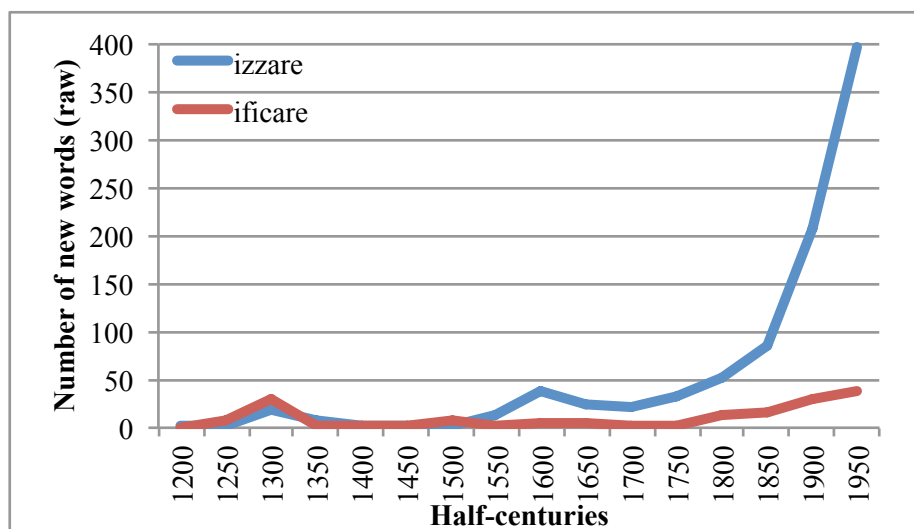


Figure 4: emergence of productivity of *-izzare* and *-ificare* in Italian.

In Figure 4, we see the raw Italian data, showing an initial increase in *-izzare* words in the 16th Century, and a surge in productivity around the 19th Century.

Synchronically, Italian appears to be similar to the other Romance languages and English:

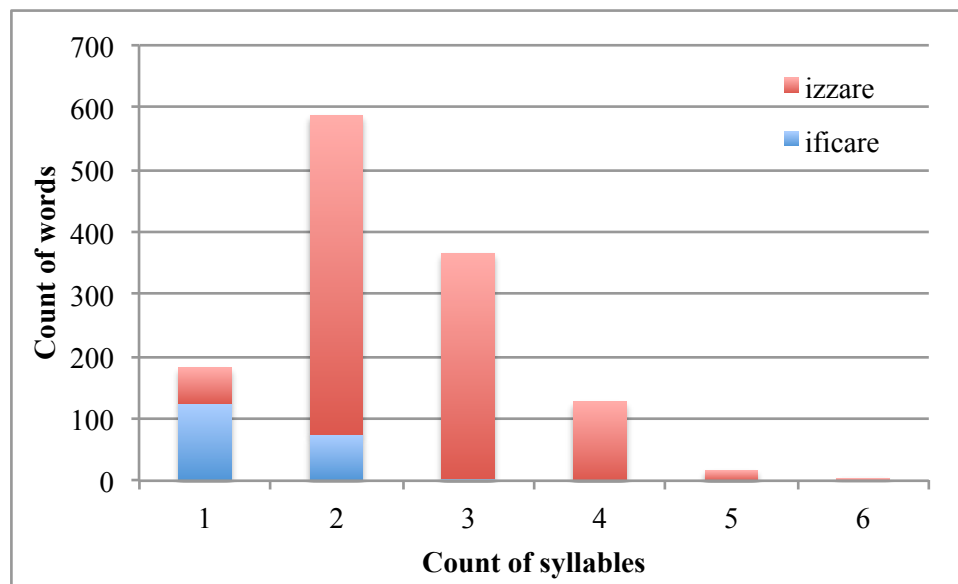


Figure 5: distribution of *-izzare* and *-ificare* words in Italian.

Fabio Montermini systematically counted the number of stem syllables in all *-izzare/-ificare* words in the electronic dictionary, with results in Figure 5. We see *-ificare* words most strongly in words with monosyllabic stems, and *-izzare* words with polysyllabic stems.

English *-ization* and *-ification*

Lindsay and Aronoff (2012) extended an investigation from Anshen and Aronoff (1999) that compared suffixes *-ment* and *-ity*; while *-ity* continues to be productive, *-ment* does not. Figure 6 below compiles historical data from the Oxford English Dictionary (OED) to show the number of new derived forms of *-ity* and *-ment* occurred over the past 800 years. We see that *-ity* and *-ment* were both building in productivity until around the 1600s when their trajectories diverged; *-ment* faded out of use.

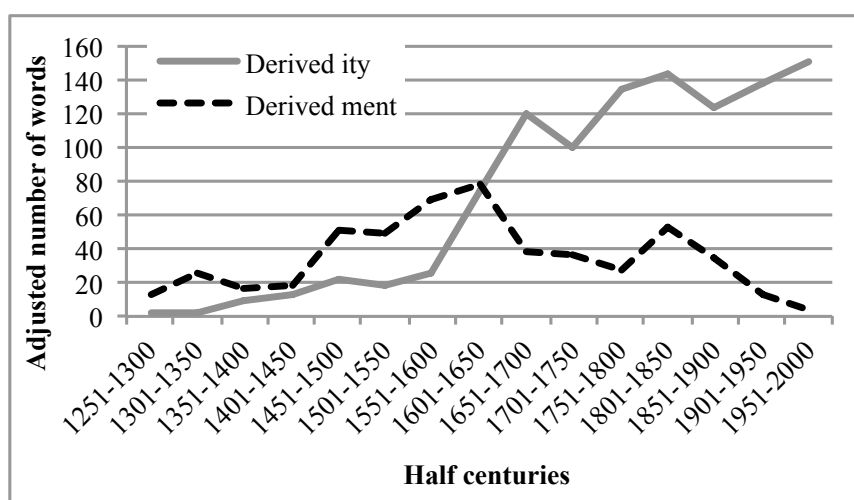


Figure 6: derived *-ity* vs. derived *-ment* (adjusted¹).

¹ The number of words has been adjusted to account for the varying number of words recorded in the OED over this span of time; the value for the number of words in a given half-century is proportional to the total words in the OED for that time period:

Lindsay and Aronoff (2012) offered two contributing factors for *-ment's* productive demise.

First, there was a sudden, significant reduction (one-third during the 1600s) in the number of new verbs entering English around the same time period, as illustrated below:

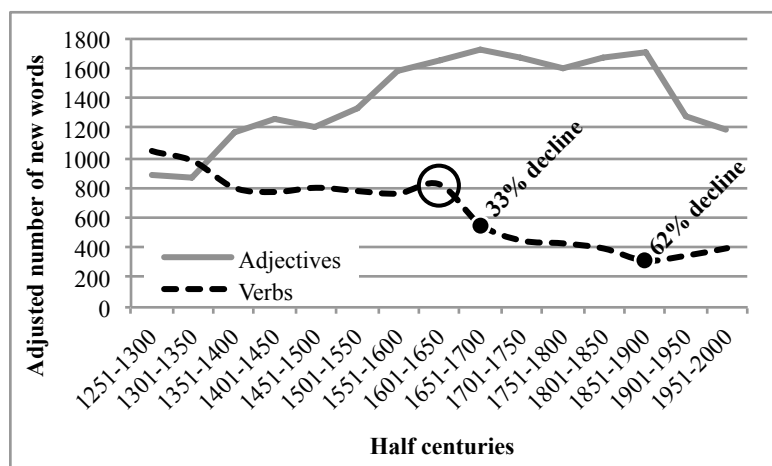


Figure 7: total new adjectives vs. new verbs per half-century (adjusted).

Since *-ment* is a deverbal suffix, its source of productivity was greatly reduced, while *-ity's* source, adjectives, was not.

Second, *-ment* had to contend with a superior rival suffixation pattern: *-ation*. English was borrowing in words from French and Latin containing *-ment* and *-ation* in addition to the derived forms (indeed, these whole-word borrowings gave rise to the suffixes' productivity). However, as we see in Figure 8, during this critical time period in the 1600s when new verbs were scarcer, borrowings containing *-ation* outnumbered *-ment* borrowings 5 to 1:



Figure 8: borrowed *-ation* vs. *-ment* per half-century (adjusted).

The high number of borrowings helped to promote the *-ation* pattern over the *-ment* pattern.

In Toulouse, we found an additional, more satisfying explanation for *-ation*'s success. In a similar investigation in Lindsay and Aronoff (2012), it was found that rivals *-ic* and *-ical* can coexist because *-ical* co-occurs with stems ending in *olog* — a phenomenon known as potentiation (Williams 1981). While *-ic* is preferred overall by an 8:1 ratio, *-ical* is more productive than *-ic* (by a similar ratio) in these *olog* stems. A similar form of potentiation was uncovered that not only reinforced *-ation*, but precluded derivation from *-ment*.

Since both *-ment* and *-ation* arose productively in English due to whole-word borrowings from French and Latin (Marchand 1969), it is possible that further subpatterns in *-ment* and *-ation* words that existed in these languages might also have carried over into English. For example, the deadjectival suffix *-ity*, which was borrowed through whole words from French, often occurred adjacent to French adjectival suffixes, like *-able*. As a result, the subpattern *-ability* was prevalent in these borrowed words from which English speakers generalized a pattern, leading to *-ity* being highly productive in new words ending in *-able*.

In French, suffixes *-iser* and *-ifier* are nominalized in a specific way, becoming *-isation* and *-ification*, respectively. While French also has deverbal suffix *-ment*, it is never an option when deverbalizing words ending in *-iser* and *-ifier*. During the centuries in which *-ment* and *-ation* words were entering English, this dichotomy would have been reflected in the structure of the words; there were no words of the form *-izement* or *-ifyment*, but there was positive reinforcement of *-isation* and *-ification*. We measured the impact of this interaction quantitatively by examining *-ization* and *-ification* words in the OED Online.

We found that, indeed, a large number of *-ation* words are of the form *-ization* or *-ification*. Out of 2555 *-ation* words in the OED, 538 are *-ization/-ification*: 21.1% of the total. Thus, the co-occurrence of *-ation* with these suffixes was not only absolute, but it comprised a rather significant number of words.

The timing of the influx of *-ization/-ification* words into the language was also informative. These forms began to rapidly increase in number around the same time that *-ment* began its decline, and continued to the present day. In Figure 9, we see the percentage of *-ation* words that were *-ization/-ification* forms.

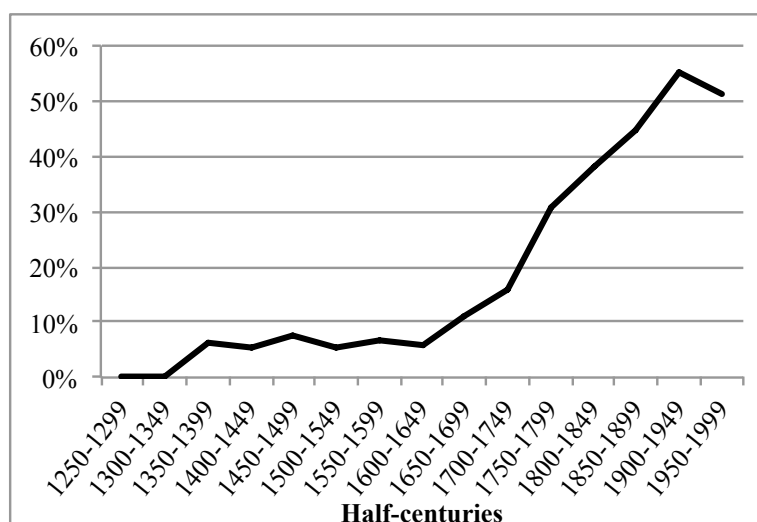


Figure 9: percentage of *-ation* words entering English that were of the form *-ization* or *-ification*.

While only 5.8% of *-ation* words were *-ization/-ification* forms before 1600, this number jumped to 30.9% for *-ation* words after 1600. This rise in *-ization/-ification* forms also follows the rise in productivity of *-ize* and *-ify*, which themselves became highly productive in the 1500s (particularly *-ize*). Thus, the *-ization/-ification* pattern had yet another avenue for propagation beyond *-ize* and *-ify* borrowings: *-ize* and *-ify* derivations. This is reflected in the *-ization/-ification* data, as 79.3% of all *-ization/-ification* forms are derived in English.

Future collaboration

Fabio Montermini and I are intending to continue the work on the Italian historical data, comparing the diachronic development of *-izzare* and *-ificare* in Italian to that of French and English

In addition, Jesse Tseng and I plan to collaborate later this summer with Daniel Huber on his work on the relationship of stress in English to the productivity of *-ize* and *-ify*, and the correlation of syllables to the selection of these forms in French, using data I have extracted from Google and the OED along with data from Tseng and Huber.

References

- Anshen, Frank, Mark Aronoff. 1999. "Using dictionaries to study the mental lexicon". *Brain and Language* 68.10.
- Lindsay, Mark. 2012. Rival suffixes: synonymy, competition, and the emergence of productivity. To appear in *Proceedings of the 8th Mediterranean Morphology Meeting*.
- Lindsay, Mark and Mark Aronoff. 2012. Natural Selection in Self-Organizing Morphological Systems. To appear in Fabio Montermini, Gilles Boyé, and Jesse Tseng (ed.) *Selected Proceedings of the 7th Décembrettes*. Germany: Lincom Europa.
- Marchand, Hans. 1969. *The categories and types of present-day English word-formation. A synchronic-diachronic approach*. Munich, Germany: Beck.
- Williams, Edwin. 1981. "Argument structure and Morphology". *Linguistic Review* 1: 81-114.