Rival suffixes:
synonymy, competition, and the emergence of productivity

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

1. Introduction

Complex systems can arise (and change) from the sum of numerous smaller interactions through natural selection. The term and idea are Darwin's but have broad application beyond biology. Three essential elements, when present in a continuous system, lead to an emergent process similar to natural selection:

(a) random perturbation of the system,
(b) a means of propagating some patterns, and
(c) the extinction of other patterns from the system.

All of these elements are present in derivational suffix systems. Speech errors, neologisms, diachronic phonological changes, and borrowings introduce random changes into the system. Productive derivation and exemplars ensure that successful patterns are more likely in the future. Finally, an intolerance of synonymy means that a suffix that does not continually attach to new words will ultimately cease to be productive and will therefore be eliminated from the system. (For an extended discussion of this theory, see Lindsay and Aronoff 2012.)

Although one affix will tend to dominate a broad domain, a language can settle into a stable system that includes the less competitive affixes as productive elements. This is achieved if the less-productive affix happens to find a niche: a clearly defined subdomain within its potential domain — a subsystem that is therefore distinct and predictable to a speaker in spite of a general trend towards another affix. Furthermore, in order for an affix to remain productive, this subdomain must also be a large enough subset of all eligible words that speakers can generalize its usage and that the affix will have an ongoing inflow of new words to combine with. A subdomain whose members share some linguistically salient characteristic should have the highest likelihood of establishing itself in the grammar of speakers; this characteristic (or combination of characteristics) is what is used to establish a pattern for forming new words. These characteristics could, then, be phonological, morphological, semantic, pragmatic, and so on.

The investigations in this paper build upon work from Lindsay and Aronoff (2012). I will investigate evidence for emergence in suffix systems: when new potential suffixes enter the system, the system ultimately organizes these elements; as the language evolves, the organization of elements adapts to these changes. These emergent processes in the English suffix system are a part of glossogenetic evolution (Hurford 1990, also discussed in Steels 1997 and Fitch 2010, among others), a concept that is distinct from phylogenetic evolution, i.e. the evolution of language ability in humans. In particular, I will examine rival suffix pairs *-ic*/*-ical* and *-ity*/*-ness* in English, showing that each suffix thrives because of adaptation to the system. I will then look at the pair of *-ize*/*-ify* and their equivalents in three Romance languages: French (the primary source of the English suffix), Spanish, and Portuguese. We will see how a coincidence in Latin ultimately led to the organizational system we see in all of these languages today.

2. Rival English suffixes *-ic* and *-ical*: comparison with a traditional corpus

The suffixes *-ic* and *-ical* are borrowed suffixes originating in Greek and Latin. While *-ic* directly corresponds to Greek *-ikós*, *-ical* is an etymologically redundant affix created from a combination of factors: Greek *-ikós*, Old French *-al*, and the many scientific nouns ending in *-ic* or *-ics* borrowed from Latin (Marchand 1969:236, 241). Both *-ic* and *-ical* are healthy, productive suffixes in the present day, although they are in direct competition. Lindsay and Aronoff (2012) found that these suffixes could both succeed because each suffix had developed a productive niche.

To evaluate each of these competing suffixes, productivity was measured in a novel way, by incorporating Google search results[1]: the exact literal string for words is queried, and the Estimated Total Hits (ETM) result from Google is recorded for each word; one can then look for numerical patterns in these numbers to learn about productivity.[2] By comparing ETM values for each form for a given stem (e.g. *biolog-ic* and *biolog-ical*), the more productive suffix will tend, over a large number of comparisons, to have a higher ETM value more often than the less productive suffix.

When the ETM values for each *-ic*/*-ical* pair were compared, *-ic* was found to be the "winner" in 10613 out of 11966 pairs.

	Total Stems	**Ratio**	***olog* Stems**	**Ratio**
Favoring *-ic*	**10613**	**7.84**	74	1
Favoring *-ical*	1353	1	**401**	**6.42**
Total	11966		475	

Table 1: *-ical* is productive in stems ending in *olog* (from Lindsay and Aronoff 2012)

While *-ic* was preferred overall, it was not preferred in all domains. After systematically examining all possible neighborhoods, it was revealed that stems ending in *olog* overwhelmingly favored *-ical* over *-ic*. This represents a morphologically bound niche for *-ical*.

Although the results from Google deal with data that is untagged and often noisy, we find similar results in a more traditional corpus, such as the Corpus of Contemporary American English (COCA), a balanced corpus containing 425 million words of text that are tagged for part of speech. In this follow-up study, a similar comparison of *-ic*/*-ical* rivals was conducted using word frequencies from among the 60,000 most frequent words in COCA.

[1] Other measures of morphological productivity exist, such as those discussed in Baayen (1993), Plag (1999) and Bauer (2001). The approach used in this article is not meant to replace currently existing methods; rather, it is an additional means of measuring productivity that exploits the vast amount of linguistic information contained within the World Wide Web.

[2] One must be cautious when incorporating Google ETM values into a measurement of usage. While Google is a vast and freely-available resource, it is also "noisy"; that is, individual results contain false positives due to typos, non-native speech, spam, the lack of part-of-speech tagging, and so on. Furthermore, ETM results represent the number of pages a string is estimated to appear in, not the number of occurrences. (Other discussion of such considerations can be found in Hathout and Tanguy 2002, among others.) For these reasons, it is important that little weight is placed upon the actual raw numbers themselves (only relative differences should be considered) or upon any individual word pairs. For the time being, it is also important to restrict investigations to single words, rather than phrases, due to the algorithm by which Google estimates phrasal results. A broad investigation of suffixes mitigates many of these concerns, as we are dealing with single words, regular inflection patterns, and a large number of stems.

All adjectives ending in *-ic* and *-ical* were queried in COCA; there were 1465 *-ic*/*-ical* pairs in all — a much smaller sample than the 11966 queried through Google. The number of tokens for each form was compared, and the form with a greater number of tokens was the "winner" for that pair. In total, 1197 pairs favored *-ic* over *-ical* (a ratio of 4.5 to 1). As in the previous study, in the *olog* subset, *-ical* was favored over *-ic* in 73 of 86 pairs (a ratio of 5.6 to 1).

	Total Stems	**Ratio**	***olog* Stems**	**Ratio**
Favoring *-ic*	**1197**	**4.5**	13	1
Favoring *-ical*	268	1	**73**	**5.6**
Total	1465		86	

Table 2: COCA results for *-ic* and *-ical*

As COCA is much smaller than what is indexed by Google, often only one form in each pair was present in the corpus; only 6% of pairs had both *-ic* and *-ical* forms with non-zero frequencies in COCA. Nonetheless, the results from COCA are very similar to those from Google, and lead to the same conclusions about the productivity of *-ic* and *-ical*. That is, we see clear evidence of a morphological niche in *olog* stems where *-ical* prospers.

3. Morphologically constrained *-ity* vs. *-ness*

In contrast to *-ic* and *-ical*, the origins of the suffixes *-ity* and *-ness* differ greatly from one another. Native suffix *-ness* had existed well before any *-ity* words had entered into English and was productive with both native and borrowed words; indeed, derivation from French adjectives was common by 1300 (Marchand 1969:335). The earliest *-ity* words were whole-word loans from French starting around the 14th century. As *-ity* words began entering the language, the edges of the domain of *-ness* were gradually eroded as *-ity* established a niche in which it could be productive. The establishment of a productive *-ity* follows a typical pattern for borrowed suffixes, as illustrated in Figure 1. First, whole words were borrowed that happened to end in *-ity*; over time, as these *-ity* borrowings accumulated, derived forms containing *-ity* began to emerge with increasing frequency. After productivity reached a critical mass, it continued to increase in spite of a decrease in whole-word borrowings.

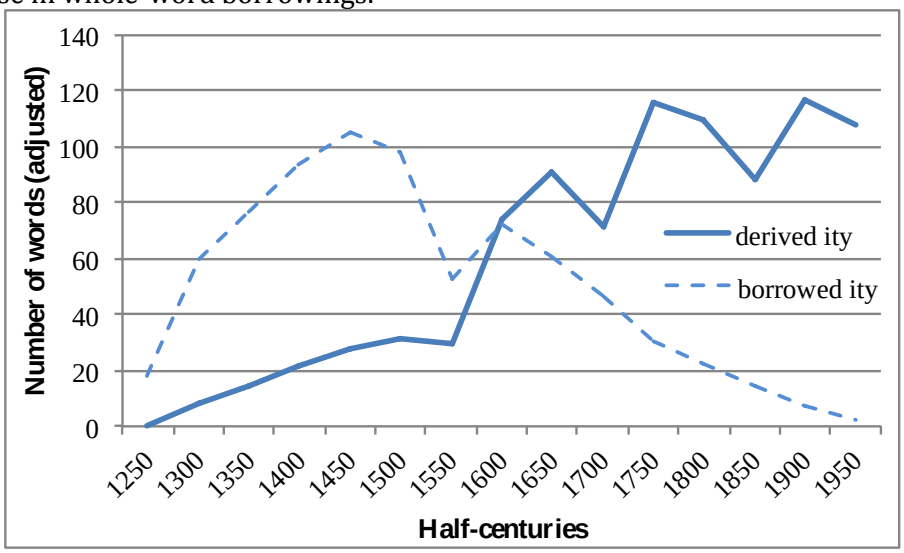

Figure 1: Whole-word borrowings of *-ity* precede its gradual development of productivity (data from first citations in the OED).

The synchronic productivity of these suffixes was compared using the methodology discussed in Section 2, which leverages Google indexing. In this case, the Estimated Total Matches were queried for 3256 rival pairs. Overall, *-ness* was found to be preferred in 2381 pairs, versus 875 for *-ity*, for a ratio of 2.72 in favor of *-ness*.
While *-ness* is more generally successful, there are a number of subdomains in which *-ity* dominates. In Table 3 below, we see the dominant subdomains for *-ity* (with at least 50 members):

	-ity	***-ness***	***ratio***
able	**1025**	379	2.7
al	**500**	183	2.7
- ial	***113***	*18*	*6.3*
- ual	***55***	*6*	*9.2*
- ral	***50***	*14*	*3.6*
ic	**181**	37	4.9
ar	**111**	18	6.2

Table 3: *-ity*/*-ness* neighborhoods favoring *-ity*

In contrast to *-ical* (which dominated only stems ending in *olog*), there are multiple large domains in which *-ity* is highly successful, each corresponding to a Latinate suffix: *-able*, *-al*, *-ic*, or *-ar*. However, its dominance is far from total: there are hundreds of forms among these sets that prefer *-ness*, a reflection of *-ness*'s overall productivity in English.

One might initially be inclined to attribute *-ity*'s preference for exclusively Latinate suffixes to some type of feature or grammatical constraint in English (e.g. Aronoff 1976, Booij 1977, Plag 1999, and others). But if we consider how *-ity* entered into English and how its place in the language evolved, such an explicit stipulation is not necessary. Because *-ity* came into English via whole-word borrowings from French and Latin, it is natural that a large number of cases would exist in which *-ity* is part of a word that already has a Latinate suffix embedded in it. In fact, it is likely because of these co-occurrences that *-ity* was able to achieve the level of productivity that it enjoys today. With native suffix *-ness* already highly productive in English, *-ity* needed to occupy a niche in order to thrive. Being part of an emergent system, *-ity* capitalized on what it was given: dominance among words containing Latinate suffixes like *-able*, *-al*, *-ic*, and *-ar*. Its productivity does not extend to any Germanic suffixes because *-ness* is already dominant there (as well as being generally dominant), thus leaving no opportunity for *-ity* to "get its foot in the door", as it were. In addition, the fact that *-ity* is also coupled with a shift in stress created an additional phonological hurdle to creeping into Germanic domains.

Nonetheless, when a suffix that hosts *-ity* has itself become productive with all bases, as in the case of *-able*, it does precisely that (Marchand 1969:313):

> With **-able/-ability** […] the synchronic relevancy of the pattern has gone far beyond its original morphological basis. While other adjectives derive substantives in **-ity** only when the adjective is Latin coined […], the derivative range of **-able/-ability** […] today comprises practically any adjective in **-able**, including adjectives derived from native roots (*lovable*/*lovability*).

Those neighborhoods where *-ness* is preferred tend to have a very different character from those that prefer *-ity*. The following are particularly noteworthy neighborhoods where *-ness* is preferred:

	-ity	*-ness*	*ratio*
ed	2[3]	**804**	
ing	0	**347**	
ess	4[3]	**281**	
ish	0	**251**	
ful	0	**250**	
ent	6[3]	**84**	
ant	3[3]	**56**	
ous/os	187	**646**	3.5
ive	128	**333**	2.6

Table 4: *-ity*/*-ness* neighborhoods favoring *-ness*

In the first seven neighborhoods above, the domination of *-ness* is total; there are effectively no exceptions (see description of false positives in footnote 3). These neighborhoods also largely correspond to Germanic suffixes: *-ed*, *-ing*, *-ess*, *-ish*, and *-ful*. The exceptions to that are *-ent* and *-ant*, which only seem to prefer *-ness*; most *-ent* and *-ant* words prefer neither suffix, but rather take *-ence/-ency* and *-ance/-ancy*, respectively. As there would not be any French borrowings ending in *-entity* or *-antity*, there would be no analogical basis for deriving such forms.

The two neighborhoods in which there is some amount of contention are *ous/os* and *ive*. In both cases, these are Latinate suffixes, coming from Old French and French/Latin, respectively (Marchand 1969:315, 339), so we might expect *-ity* to dominate. Instead, it is *-ness* that is dominant. This further underscores how a natural system evolves: just because *-ity* was able to take advantage of exemplars favoring *-ability,* and other Latinate suffixes, this does not guarantee that it will have the same level of success in all such cases. The fact that *-ity* does provide healthy competition in both *ous*/*os* and *ive* neighborhoods (in contrast to the Germanic neighborhoods) shows that its Latinate origins did have some impact on the productivity of the rivals in those domains; that is, there were some whole-word borrowings ending in *ivity* (e.g. *festivity*, *captivity*, *activity*) and *osity* (e.g. *curiosity*, *virtuosity*, *generosity*) trickling into English, providing some basis for extending the pattern, but not enough to dominate these niches.

4. Phonologically defined niche: *-ize* versus *-ify*

The English suffix *-ize* originated as Greek *-izō*, entering into English through Late and Ecclesiastical Latin *-izare* and French *-iser*. Its rival, *-ify*, comes ultimately from Latin *-ificare*, though many words came into English via French *-ifier* (Marchand 1969). Lindsay and Aronoff (2012) examined *-ize* and *-ify* in present-day English. They queried 2636 *-ize*/*-ify* suffix pairs and found that 2217 favored *-ize* and 419 favored *-ify*: a 5.0 ratio. Therefore, it would seem that *-ize* has been more productive overall. However, like *-ical*, *-ify* is productive in a certain subdomain of stems. As we see in Figure 2, the productivity of each affix correlates strongly with the number of syllables in the stem.

[3] Forms favoring *-ity* in these neighborhoods were as follows:

ed: *rubedity*, *heredity*

ess: *necessity*, *nonnecessity*, *unnecessity*, and *supernecessity*

ent: *entity*, *nonentity*, *identity*, *nonidentity*, *coidentity*, *stringentity* (which owes its high frequency to an object in the Java programming language called *StringEntity*)

ant: *quantity*, *overquantity*, *disquantity*

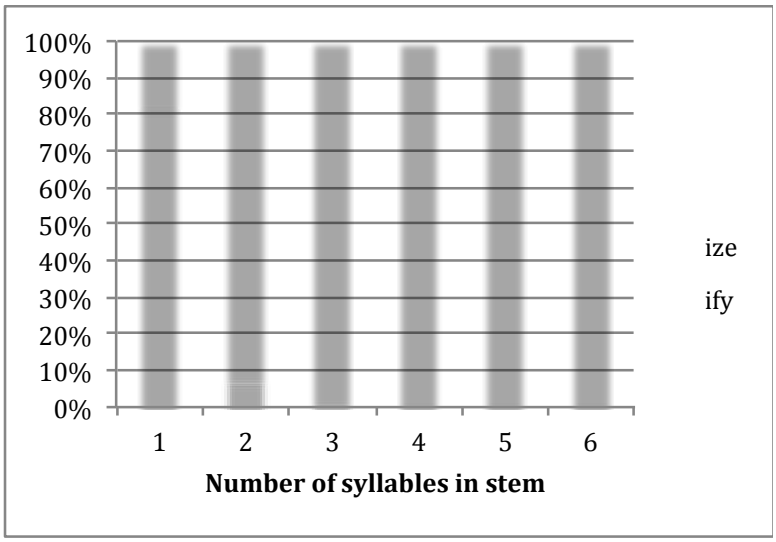

Figure 2: Strong preference for *-ify* in monosyllabic stems, sharply dropping off as stem length increases (from Lindsay and Aronoff 2012, Figure 11).

Monosyllabic stems favored *-ify* over *-ize* in 82% of comparisons, dropping sharply to just 4% for disyllabic stems. Thus, we see a strong tendency but not a strict rule: the shorter the stem, the more likely *-ify* is preferred — but neither suffix is totally restricted.

Both *-ize* and *-ify* are borrowed suffixes, which emerged out of little more than whole-word Romance borrowings that were continually entering into the language at the time. As we see in Figure 3 below, both *-ize* and *-ify* were first borrowed beginning in the late 13th century. The borrowing continued and, over time, this growing collection of borrowings began to develop organization. Both emerged as productive suffixes at the same time as well; namely, in the 16th century. From the beginning, *-ize* was the more productive of the two.

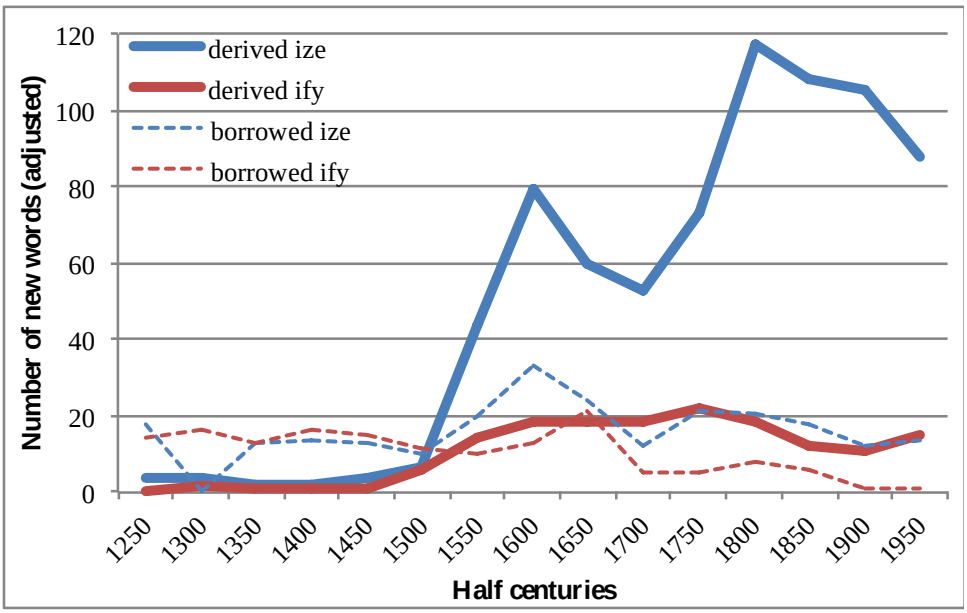

Figure 3: Derived and borrowed forms of -ize and -ify

An interesting question arises about the relative productivity of these rival suffixes in the Romance languages themselves: are the productive domains in these languages similar to English? In this investigation, we will look at the productive domains of *-ize*/*-ify* suffix

rivals in French, Spanish, and Portuguese[4], using the same method of investigation that was used for English.

A total of 365 pairs were queried in French. As with English, the full inflectional paradigm was taken into account, and the query results for all inflectional forms for a given word were summed together (e.g. 365 pairs, 37 inflections, leading to 27010 Google queries). Fortunately, the morphological paradigm for these suffixes is regular, thus making the process of generating forms much less complicated.

The overall results for French showed a similarity to those for English: *-iser* was heavily favored over *-ifier*, where it was the winner of 292 pairs: a 4.0 ratio.
In Table 5, we can see a breakdown of the number of winners grouped by syllables in the stem:

	1	**2**	**3**	**4**	**5**	**6**	**Total**
-ifier	50	22	1	0	0	0	*73*
-iser	25	144	85	32	3	3	*292*
	75	*166*	*86*	*32*	*3*	*3*	*365*

Table 5: count of types preferring each suffix (1-6 syllables in stem) for French

Like English, there is a tendency for *-ifier* to have a monosyllabic stem, as 68% of all *-ifier* winners have monosyllabic stems, while *-iser* is preferred in polysyllabic stems, where 83% of all *-iser* words occur. Indeed, we see a similar pattern to English in Figure 4:

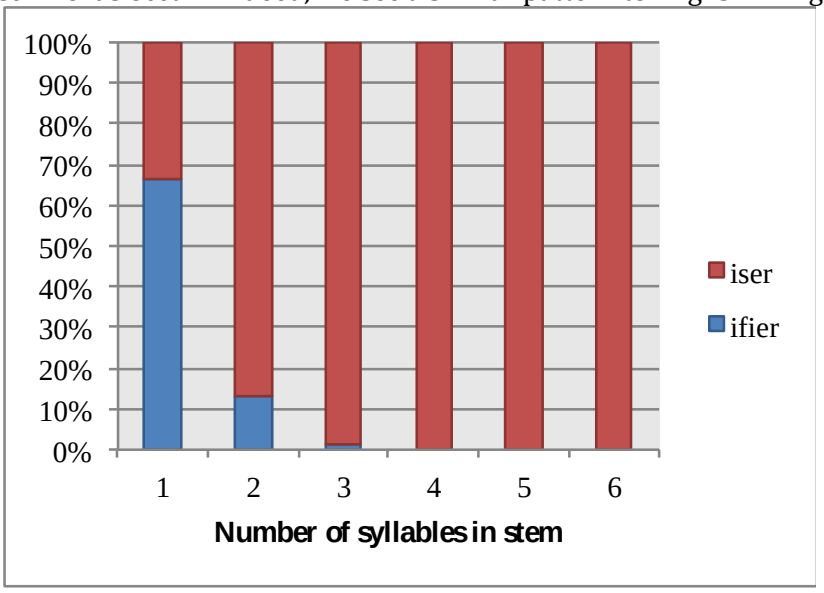

Figure 4: *-ifier* maps to monosyllabic stems, but more weakly than English

Like English, we see an attraction towards shorter stems in *-ifier*; however, *-ifier* is the winner only about two-thirds of the time with monosyllables. That is, the pattern is somewhat weaker in French than in English, but clearly present.
Next, looking at Spanish, we see that *-ificar* is clearly strongest among monosyllabic stems, with just over 50% preferring *-ificar*, while *-izar* peaks at disyllabic stems.

[4] Dictionary lists of query words were compiled largely from free dictionaries, particularly those found at *WinEdt.org*. An initial investigation into Italian was conducted as well, but was not included due to a small sample size used (100 words). However, the results did match with the other Romance languages investigated here.

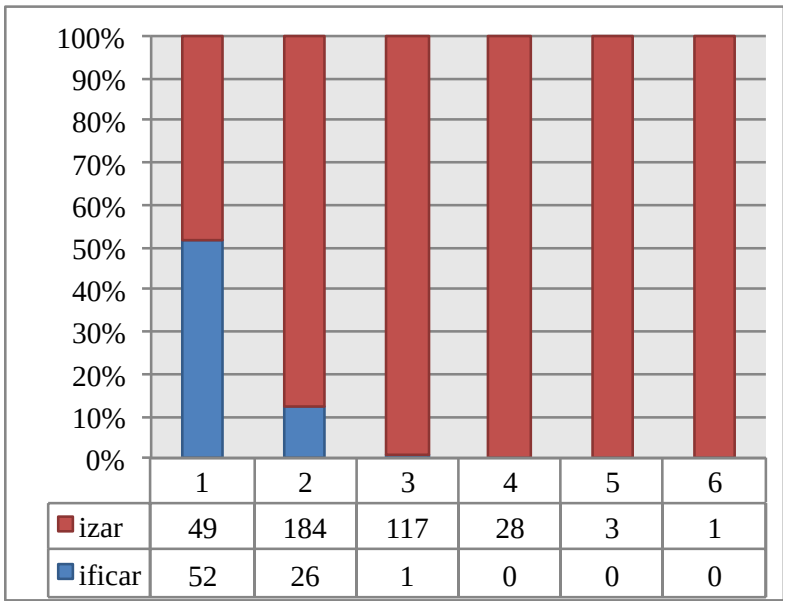

Figure 5: Spanish *-ificar* also tends toward monosyllabic stems

Portuguese also follows the pattern of its sister languages, with a monosyllabic *-ificar* tendency.

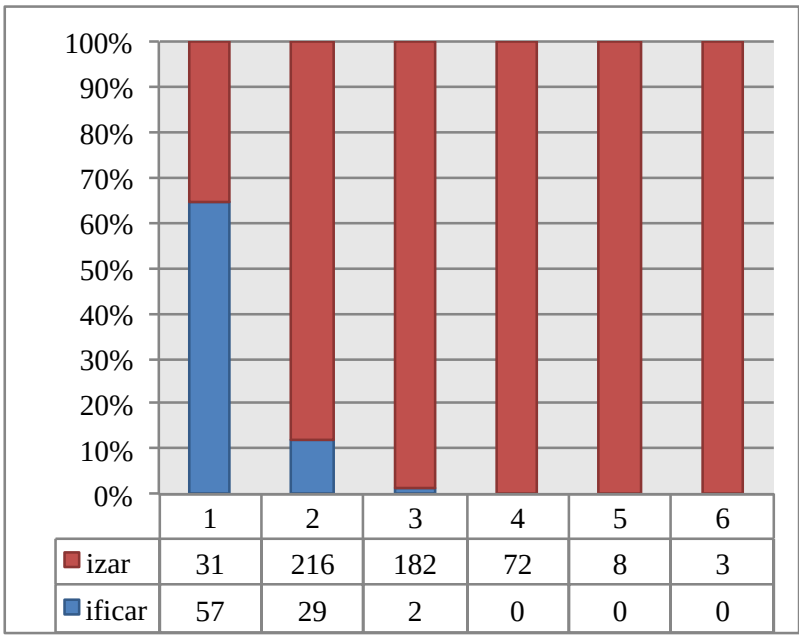

Figure 6: Portuguese shows an even stronger *-ificar* monosyllable correlation

We see a striking similarity in the relative distributions of *-ize* and *-ify* forms across all four languages. In Figure 7 below, we see the distributions of *-ize* for all four languages.

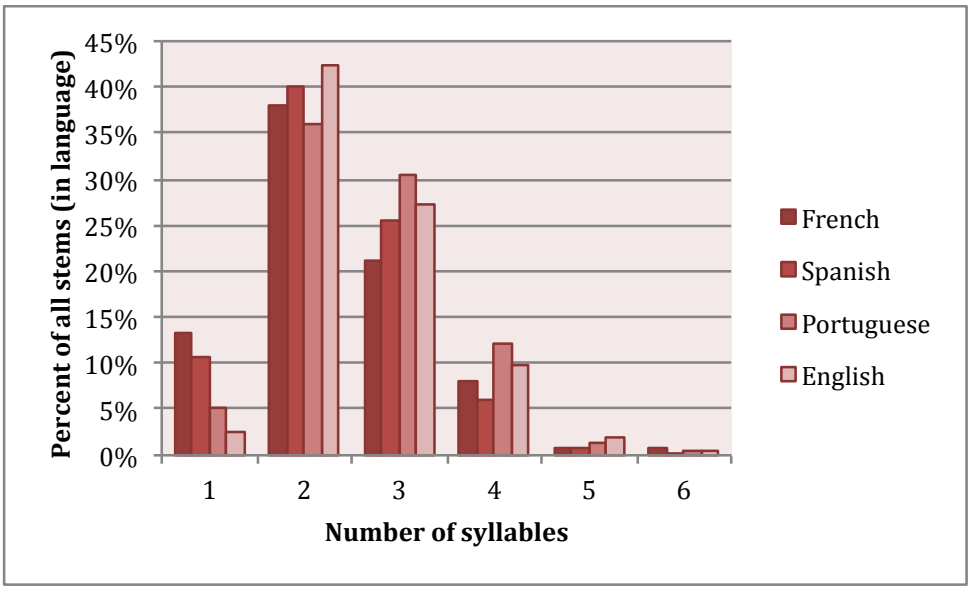

Figure 7: Relative distributions of *-ize* across all languages

In all cases, *-ize* peaks at two syllables and slowly tapers off, whereas *-ify* peaks at one syllable and drops off more rapidly, as we see in Figure 8.

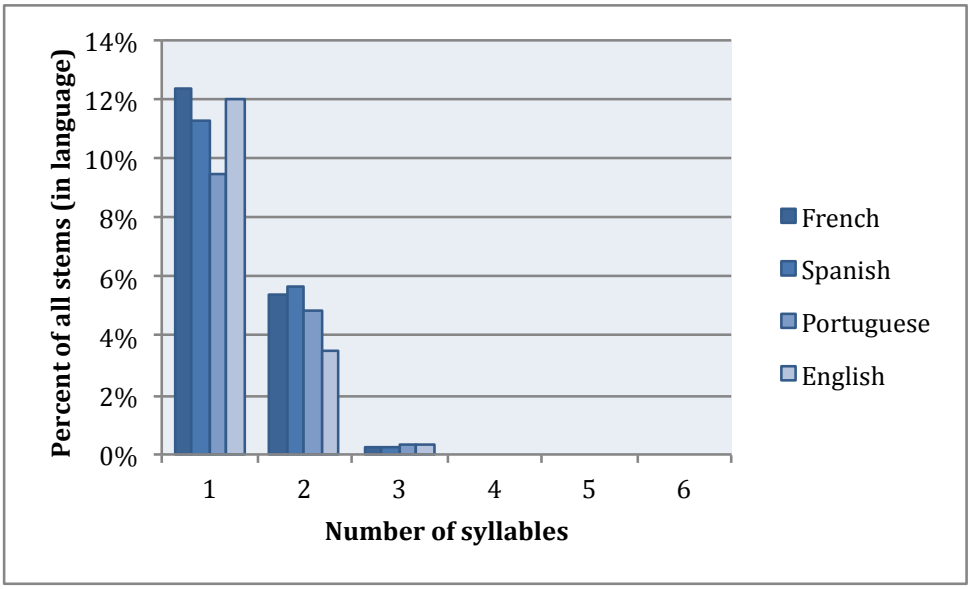

Figure 8: Relative distributions of *-ify* across all languages

Despite other phonological differences among these languages, this relationship between syllables and these suffixes is universal in the languages examined. In the case of French, Spanish, and Portuguese, this relationship was inherited directly from Latin; in the case of English, the relationship was actually recreated due to the exemplars provided by the whole-word borrowings.

We might be able to explain the source of this split, then, by looking at Latin, where both of these suffixes began[5]. Both *-izare* and *-ificare* were latecomers to Latin, with *-izare* being part of whole-word borrowings from Greek and *-ificare* being a grammaticalization of the Latin verb *facere* ('to make/do'). The latter formed a few verbs

[5] Thanks go to Mark Aronoff for suggesting this route of investigation.

in Classical Latin, more in Late Latin, and grew exceedingly frequent in Medieval Latin (Marchand 1969:300). Words containing *-izare* finally entered Latin from Greek around the 3rd or 4th century (Marchard 1969:318) with the conversion of the Roman Empire to Christianity. As such, most Latin *-izare* words have Greek or Semitic stems (91.9%) and are largely ecclesiastical in nature (e.g. *euangelizare*, *hymnizare*, *prophetizare*). In contrast, *-ificare* words almost exclusively contain Latin stems (98.8%). This divide between Latin and Greek/other turns out to be crucial for the patterning of these suffixes in all of the modern languages that have been discussed in this section. Because Latin naturally tends toward monosyllabic stems, while Greek stems tend to be polysyllabic, the number of syllables in the stems averages to 2.03 for the *-izare* words and 1.04 for *-ificare*. These results are summarized below:

	-izare	*-ificare*
Words	35	82
Latin stems	3	81
Non-Latin stems	32	1
Percent Latin stems	8.1%	98.8%
Average syllables in stem	2.03	1.04

Table 6: Summary of Latin -izare and -ificare words

We know that *-ize* and *-ify* came into English through whole-word borrowings from French. The *-iser*/*-ifier* data show that French was not only the source of two new productive English suffixes; in addition, English speakers' grammars were also influenced by the way that the forms were distributed in French. This led to a recreation in English of the productive niches for each of these suffixes (based on number of syllables) in addition to the suffixes themselves.

Furthermore, if the distribution of *-iser* and *-ifier* in present day French is similar to the distribution several centuries ago (i.e. representing a mono/bisyllabic split that is somewhat weaker than English), this would suggest that English took this pattern from French and "ran with it", not just respecting the pattern but strengthening it.

Thus, the patterning of *-izare* and *-ificare*, the equivalents in the Romance languages, and ultimately English, was the result of a coincidence: two rival verbalizing suffix patterns happened to enter into Latin around the same time; Latin natively preferred monosyllabic stems while the Greek words had longer stems. This superficial correlation between stem length (or overall word length) and suffix was recognized by Latin speakers and was salient enough to propagate into (and sustain itself in) all of the other Romance languages we have examined. Finally, and most interestingly, this pattern was even recreated in the English language when the suffixes were borrowed, based on the words English speakers were encountering.

5. Conclusion

Rival suffixes *-ic* and *-ical*, made of borrowed elements, are both productive today because *-ical* has carved out a morphologically constrained productive niche: it is preferred in stems ending in *olog*. This pattern is clear when comparative productivity is measured using data from Google, as well as a comparative sample from COCA, a more traditional corpus.

We saw the emergence of productivity in the suffix *-ity*, where whole-word French borrowings over several centuries ultimately led to a productive suffix. Because native rival *-ness* was already productive in English from the start, *-ity* was only able to emerge because of a high co-occurrence in several common Latinate neighborhoods (including *able*, *al*, *ar*, and *ic*), allowing a niche pattern to enter the system. This pattern of co-

occurrence continued to be the basis for deriving new *-ity* forms, even as suffix *-able* began to extend its productivity beyond Latin stems.

Lastly, while *-ize* is more productive than *-ify* overall, *-ify* is preferred overwhelmingly in words with monosyllabic stems. English, as well as other Romance languages, owes this particular pattern to Latin. Its native suffix *-ificare*, grandfather of *-ify*, developed through grammaticalization, and as such, conformed to Latin patterns of affixation that preferred monosyllabic stems. When Latin later began borrowing *-izare* words from Greek, these words contained naturally longer Greek stems. Though accidental, this distribution was recognized as a pattern by speakers and formed the basis of the productive niches of these two suffixes, which propagated into Spanish, Portuguese, and French (and probably others). Later, like *-ity*, *-ic*, and other borrowed suffixes, *-ize* and *-ify* entered English via whole-word borrowings, in this case from French. Since both *-ize* and *-ify* words were being borrowed into English at the same time, the same pattern was recognized by English speakers and recreated as these suffixes emerged simultaneously as productive affixes several centuries later.

In each of these rivalries, we see the organization of new elements entering the language. Suffix *-ical* emerged from the conflation of two other productive suffixes, and remains productive despite its redundancy because it found a niche in one large subset of stems. Meanwhile, *-ity* became productive in a system that already had a highly productive native suffix, only due to the influx of a preponderance of superficially similar words ending in *-ity* that eventually could not be overlooked. The productive pattern of *-ize* and *-ify* reflects a superficial prosodic difference between Latin and Greek that was many languages (and many centuries) removed by the time the very same derivational pattern emerged anew in English. Since language is an evolving system, this organization emerges gradually out of whatever bits and pieces the system happens to have in front of it.

References

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

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.

Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

Booij, Geert. 1977. *Dutch morphology: A study of word formation in generative grammar.* Holland: Foris.

Fitch, W. Tecumseh. 2010. *The Evolution of Language.* New York: Cambridge University Press.

Hurford, J. 1990. Nativist and functional explanations in language acquisition. In I. M. Roca (ed.), *Logical Issues in Language Acquisition*. Dordrecht: Foris Publications, 85-136.

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.

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

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