

Morphological Productivity Rankings of Complex Adjectives

Stefano Vegnaduzzo

Ask.com

Oakland, CA 94607, USA

svegnaduzzo@ask.com

Abstract

This paper investigates a little-studied class of adjectives that we refer to as ‘complex adjectives’, i.e., operationally, adjectives constituted of at least two word tokens separated by a hyphen. We study the properties of these adjectives using two very large text collections: a portion of Wikipedia and a Web corpus. We consider three corpus-based measures of morphological productivity, and we investigate how productivity rankings based on them correlate with each other under different conditions, thus providing different angles both on the morphological productivity of complex adjectives, and on the productivity measures themselves.

1 Introduction

Adjectives as a syntactic category have received relatively scarce attention in theoretical and computational linguistics, at least when compared to noun and verbs. Within the class of adjectives there is a particular group that has not received much specific attention. These are the adjectives that can be operationally defined by the fact that in the orthography they are constituted by at least two tokens separated by a hyphen. Examples include: *left-wing politician*, *best-selling book*, *part-time job*, *large-scale experiment*. In this paper we will focus the analysis on 2-item adjectives; however the class includes several examples with more than two items: *day-to-day*, *tongue-in-cheek*, *up-and-coming*, *state-of-the-art*, *pay-as-you-go*, *all-you-can-eat*, etc.

We refer to these adjectives as ‘complex adjectives’. They allow several mechanisms for the generation of new linguistic expressions which are

spread along a continuum that goes from full productivity to lexicalized forms. Treating complex adjective formation as a morphological process allows us to characterize them in the light of the notion of morphological productivity.

Morphological productivity falls within the domain of how the broad notion of creativity is realized in language, by providing mechanisms for generating new words that are *unintentional*, *unlimited* and *regular* (Evert and Lüdeling 2001).

Previous work has typically considered complex adjectives only in the context of tokenization and generally low-level text processing, without specific focus on the class *per se*. Moilanen and Pulman (2008) identify polarity markers in certain complex adjectives for the purpose for assigning sentiment polarities to unknown words (*well-built* is positive, *rat-infested* is negative).

Highly productive classes like complex adjectives are problematic for computational lexicons in that they bring about a large number of unknown words. A typology of complex adjectives is important to support computational lexicons and the NLP applications based on them, for example by listing out the patterns (as defined in section 3) that complex adjectives are based on.

2 Data

The analysis presented here is based on two large text collections. The first is a portion of Wikipedia, consisting of about 250 million tokens. The second collection is a Web corpus specifically built for NLP applications targeted at adjectives, consisting of about 290 million tokens (after some reasonable clean-up process). The list of all adjectives in WordNet¹ was used as a seed list. Each seed adjective was sent as a query to the Yahoo search engine

¹ <http://wordnet.princeton.edu>.

BOSS API². For all the 1000 (the maximum allowed by the API) returned web results, each URL was used to fetch the corresponding web page, and the full text of the web page was processed for addition to the corpus. Both corpora were POS-tagged using the Stanford POS tagger³ (reported accuracy is about 97%). Using such tagger that was trained on the Penn Treebank makes it easy to identify complex adjectives, since the Penn Treebank tagging guidelines explicitly prescribe that hyphenated compounds used as modifiers should be tagged as adjectives (JJ).

The two corpora give a slightly different view on the linguistic behavior of adjectives. The Wikipedia corpus is intended to provide a picture of adjective distribution in a large and relatively homogenous collection of current English text. The Web corpus is intended to bias the collection in the direction of making sure that at least for the WordNet seed adjectives a large number of instances are present even for adjectives that would have otherwise a very low frequency in a properly balanced corpus.

Throughout the paper we will use the Wikipedia corpus as the main data collection for the presentation of the analysis, and we will use the Web corpus as a validation set, to assess the stability and reliability of the results obtained on the Wikipedia corpus.

3 Theoretical background

We partition the orthographically defined class of complex adjectives in morphological categories defined by a variable part and a base generator. For example ‘X-free’ is a morphological category that specifies that the base generator *-free* can be combined with other words to form complex adjectives such as *risk-free*, *toll-free*, *gluten-free*, etc. The two corpora are used to analyze and quantify how different degrees of linguistic creativity are exhibited by different classes of complex adjectives. We will show that this yields a ranking of complex adjective types along a continuum going from high productivity to lexicalized forms. Specifically, we characterize the linguistic creativity of complex adjectives through three distinct and complementary measures of morphological productivity, following Baayen (1993, 2006) and his notation:

realized productivity, expanding productivity, and potential productivity. Now, Evert and Lüdeling (2001) show that, in the general case, automatic pre-processing of text corpora with current morphological analyzers yields results that are too noisy for Baayen’s measure of productivity to be reliable. The fact that complex adjectives are orthographically defined by the presence of the hyphen and the components are easy to separate eliminates some of those text-processing problems. By choice, we take at face value the morphological parse provided by the hyphen, and therefore we do not run into situations where the phonotactics obscures the morphological analysis (e.g., *lady* vs. *ladies*), affix ordering (*undoable*), accidental string identity (*restaurant* does not instantiate the prefix *re-*), and words generated by creative rather than morphological productive processes (*youtube*). Relying on the orthographical hyphen eases the problems of automatic morphological pre-processing and makes this a particularly good domain for using Baayen’s measures.

Realized productivity is defined as $V = V(C, N)$, the number of word types (as opposed to word tokens) of morphological category C in a corpus of N tokens. The intuition behind this measure is that it expresses the sheer size of a morphological category within a particular corpus.

Expanding productivity is defined as $P^* = V(1, C, N) / V(1, N)$, where $V(1, C, N)$ is the number of words of category C that occur only once in a corpus of N tokens, and $V(1, N)$ is the number of words of *any* category that occur only once in a corpus of N tokens. This measure expresses the contribution of the morphological category C to the growth rate of the total vocabulary. The underlying intuition is that a morphologically productive category contributes to the growth rate of the vocabulary of a language, and vocabulary growth rate can be estimated by the number of *hapax legomena* (words with frequency 1) in a sufficiently large corpus. *Hapax legomena* in turn are taken to be good estimators of novel linguistic forms.

Potential productivity is defined as $P = V(1, C, N) / N(C)$, where $N(C)$ is the total number of tokens in the corpus for a given category C. This measure expresses the growth rate of the vocabulary of the category C itself. In other words, it expresses the ease with which a category can be applied to new words.

² <http://developer.yahoo.com/search/boss>.

³ <http://nlp.stanford.edu/software/tagger.shtml>.

These three measures are complementary, in that they capture different aspects of morphological productivity that can possibly be at odds with each other (Baayen 2006).

4 Analysis

At the highest level the two POS-tagged corpora allow us to compare the distribution of complex adjectives with respect to other major syntactic categories. In the context of morphological productivity, the *vocabulary* of a morphological process is the number of types that the process can potentially generate. If we understand complex adjective formation as a morphological process, we can obtain their general vocabulary growth curve by plotting the number of types against the number of tokens in a corpus. Figure 1 shows vocabulary growth curves for nouns, verbs, adjectives and their hyphenated counterparts for the Wikipedia corpus.

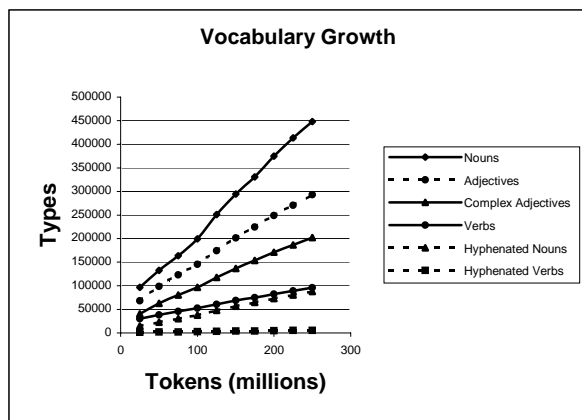


Figure 1. Wikipedia vocabulary growth.

Complex adjectives exhibit a surprisingly high growth rate, which at 250M tokens still doesn't tend to converge. The number of types corresponding to the final corpus size can be interpreted as an estimate of the realized productivity of complex adjectives as a whole class.

We focus the analysis on two-item adjectives, i.e., those of the form 'A-B' such as *cat-like*.

We describe first the data collection process. We process each POS-tagged corpus from beginning to end, and for each complex adjective (defined by the tag JJ and the presence of a hyphen), we generate all the possible categories by replacing in turn each item with a variable. So *cat-like* yields the two possible categories 'cat-X' and 'X-like'; *day-to-day* yields 'X-to-day', 'day-X-day', 'day-to-X'

and the special case 'X-to-X', where the two side items are identical. Then for each category we count the number of types in that category and the number of tokens for each type, including *hapax legomena*. So the category 'X-like' includes the types *dog-like*, *cat-like*, *mouse-like*, etc., and each type is instantiated by a certain number of tokens (including types instantiated by exactly 1 token). In this way we obtain the basic counts that are needed to compute the three measures of productivity introduced above. Once these measures are obtained, all morphological categories (such as 'X-like') can be sorted into three productivity rankings corresponding to the three measures.

Of the three measures, realized and expanding productivity are by design directly dependent on corpus size, whereas potential productivity is dependent on the relative ratio of morphological categories but not directly on corpus size. This entails that, especially for the first two measures, productivity rankings obtained on a corpus of a particular size are not necessarily significant, in that productivity rankings obtained on corpora of different sizes could be different.

Realized and expanding productivity are affected in different ways by corpus size. Everything else being equal, realized productivity estimates will be more reliable with larger corpora. Expanding productivity is based on the assumption that *hapax legomena* (i.e., rare words with the lowest frequency in a corpus) are good estimators of morphological productivity. Rare words include various subtypes: misspellings, proper names, foreign words, words from different registers or genres than those represented in the corpus, new words generated by non-regular creative processes, and new words generated by morphologically productive processes. In a 'small' corpus, *hapax legomena* will include many words that do not fall in any of categories above, but just happen to be relatively uncommon words. As corpus size increases, uncommon words have a chance to occur more often, and the proportion of true morphological neologisms among hapaxes increases.

Given the dependency on corpus size, the question arises of how reliable productivity rankings are that haven't been obtained for a specific corpus. In order to assess rankings reliability, we divided both the Wikipedia and Web corpus into 20 chunks of the same size, and then we added them up one after the other, essentially obtaining 20 corpora of

increasing size for each of the two original text collections. Finally for each intermediate corpus we recomputed the productivity measures and rankings based on them. At this point we can measure the stability of the final productivity ranking for a given corpus by comparing it with each of the intermediate rankings using Spearman’s rank correlation coefficient.

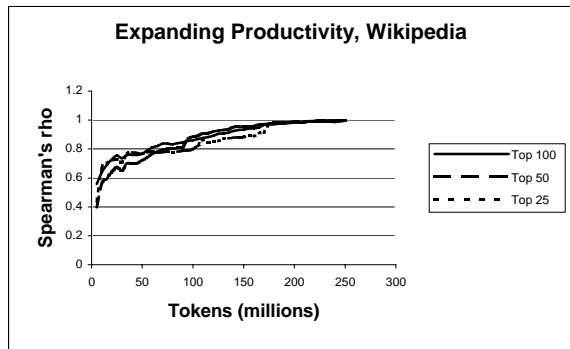


Figure 2. Expanding productivity stability, Wikipedia

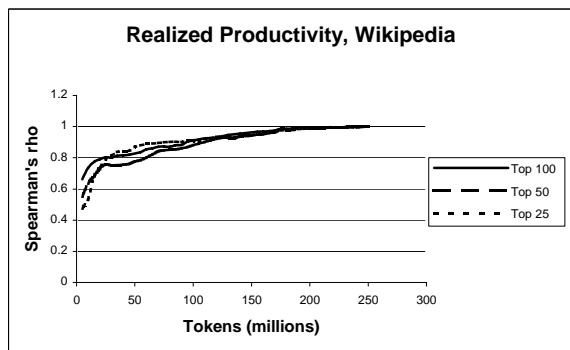


Figure 3. Realized productivity stability, Wikipedia

Plotting corpus size against Spearman’s ρ coefficient reveals that rankings for realized and expanding productivity in both the Wikipedia corpus and the Web corpus are very stable. For example, for expanding productivity in the Wikipedia corpus the ranking of the top 20 complex adjectives does not change any more after the first 5M corpus chunk. Figures 2 and 3 show how Spearman’s coefficient between the full Wikipedia corpus and each of the 19 sub-corpora varies as a function of corpus size for the top 25, 50 and 100 adjectives. Curves for the Web corpus are similar.

As corpus size continues to increase, a larger and larger portion of the rankings stabilizes, suggesting that realized and expanding productivity capture substantial properties of the language of which the Wikipedia corpus and the Web corpus are respectively representative.

The dependency of productivity rankings on a specific corpus is a reminder that another important factor that has been observed to affect linguistic creativity as expressed by morphological productivity is register and genre variation. In this context we are using two corpora that are representative of different types of language, and could therefore have substantially different rankings. However, it turns out that Spearman’s rank correlation between the full size of the Wikipedia and Web corpus for both realized and expanding productivity rankings is quite strong (see Table 1).

Rankings	Realized	Expanding
Top 5	1	0.9
Top 10	0.91515	0.94545
Top 20	0.85939	0.70601
Top 25	0.80346	0.665
Top 50	0.64998	0.58909
Top 100	0.64309	0.62732

Table 1: Spearman’s correlation between productivity rankings for Wikipedia and Web corpus

The high Spearman’s correlation values confirm the stability of reliability of realized and expanding productivity rankings across different corpora.

The next question is to what extent productivity rankings based on different measures correlate with each other. We consider first the relationship between realized and expanding productivity. The two notions focus by design on different aspects of morphological productivity: realized productivity is oriented towards the ‘present’: a morphological process may be common to many existing word types, but might have no ability to be applied to generate new words. Baayen (1993) also refers to it as ‘extent of use’. Expanding productivity is intended to assess the rate at which a morphological process is expanding in the language. A morphological process may be able to spread quickly in terms of its ability to generate new words, and yet not be very common in the general language.

In order to quantify the degree of agreement between realized and expanding productivity for complex adjectives we computed Spearman’s rank correlation coefficient between pairs of productivity rankings (for the top 100 adjectives) at each of the intermediate corpus sizes described above, for both the Wikipedia and Web corpus.

Figure 4 shows that for complex adjectives in the Wikipedia corpus (the graph is similar for the

Web corpus) the correlation between realized and expanding productivity is very strong, the value of ρ being constantly around 0.9 from the very beginning, therefore independently of corpus size. We can interpret this fact as suggesting that the class of complex adjectives is overall a dynamic class, in the sense that its members, even the most established (i.e., those with high realized productivity), continue to expand in the language and to allow speakers to easily create new words. Table 2 shows the top 20 complex adjectives for both realized and expanding productivity.

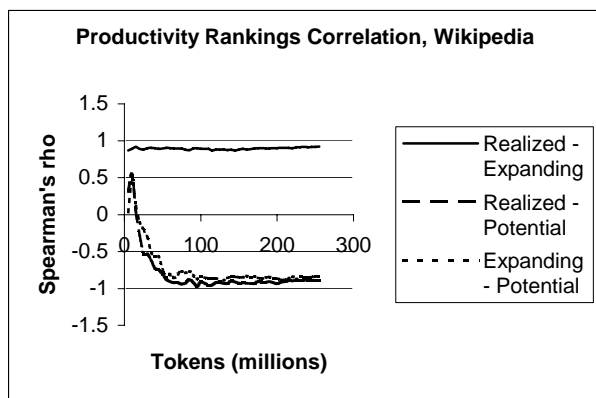


Figure 4: Productivity Rankings Correlation for the Wikipedia corpus, top 100 complex adjectives.

Figure 4 also shows that Spearman's correlation of both realized and expanding productivity with respect to potential productivity is significantly weaker. It is important to note that in the case of potential productivity we arbitrarily chose to require that the total number of tokens in the corpus for a given morphological category (the denominator of the potential productivity formula) be at least 100. We will show later that choosing a high value, (such as 100) for this factor indeed increases the strength of the correlation between potential productivity and the other two types. With low values the correlation coefficient is closer or equal to -1 .

We are now going to discuss how rankings for potential productivity (and, consequently, their degree of correlation with realized and expanding productivity) vary with the choice of the minimum threshold for the denominator value of the formula $P = V(1, C, N) / N(C)$.

In general the situation for potential productivity is quite a different. On one hand it doesn't depend directly on corpus size like the other two measures. On the other hand, its definition as the ratio of *hapax legomena* of a morphological category with

respect to all tokens of that category makes it susceptible to frequency effects. If a corpus is too small or a morphological category is fairly rare this measure will overestimate the productivity of low frequency items. In the extreme case, a morphological category instantiated by exactly one type that occurs exactly once in the corpus will get a potential productivity value of 1, the highest possible.

Wikipedia productivity rankings			
Realized		Expanding	
1.non-X	2.53E-05	1.non-X	0.014388
2.X-based	1.63E-05	2.X-based	0.009694
3.X-like	1.49E-05	3.X-like	0.00943
4.anti-X	9.59E-06	4.anti-X	0.005875
5.pre-X	8.00E-06	5.pre-X	0.005001
6.X-style	6.58E-06	6.X-style	0.004809
7.X-related	6.38E-06	7.X-related	0.004395
8.X-type	5.36E-06	8.X-type	0.004386
9.post-X	5.35E-06	9.post-X	0.003443
10.self-X	4.23E-06	10.then-X	0.002381
11.semi-X	3.78E-06	11.semi-X	0.00224
12.multi-X	3.73E-06	12.self-X	0.002197
13.re-X	3.64E-06	13.ex-X	0.002116
14.then-X	3.42E-06	14.X-oriented	0.002082
15.pro-X	3.41E-06	15.re-X	0.002078
16.X-oriented	3.22E-06	16.pro-X	0.002022
17.ex-X	2.97E-06	17.un-X	0.001937
18.single-X	2.95E-06	18.multi-X	0.001852
19.two-X	2.90E-06	19.X-only	0.001813
20.high-X	2.78E-06	20.half-X	0.001604

Table 2: Wikipedia rankings for realized and expanding productivity.

The notion of potential productivity is often used, especially in Baayen's work, in a deductive setting: typically a relatively small number of derivational morphemes is selected, and an in-depth analysis is carried out, for example assessing intuitive productivity rankings against those obtained from corpus statistics. The goal is often to achieve high explanatory depth, integrating for example mental processing and socio-linguistic factors. In this context the target morphemes are typically fairly common in the language, and a sufficiently large corpus will provide a sufficiently large sample of each morphological category so that the frequency effects mentioned above are negligible.

On the other hand, the setting of the work presented in this paper is inductive in nature, in that, by capitalizing on the easy identification of complex adjectives thanks to the orthography, we aim at an exploratory characterization of a large class of morphological categories.

In practice, in the large-scale scale setting of the present work, frequency effects have a significant consequence on how potential productivity values are computed, and on the rankings that derive from them. For the Wikipedia corpus 49.36% (40223 out of 81481) of the complex adjective morphological categories have only 1 type with 1 token. For example the category ‘X-distracted’ (based on a past participle like the much more common category ‘X-controlled’, which has types such as *computer-controlled*, *electronically-controlled*, etc.) has only the type *easily-distracted*, which occurs only once in the corpus. By blindly applying the formula $V(1,C,N)/N(C)$, we would conclude that complex adjectives that occur only once have the highest potential productivity ($1/1=1$), which runs counter to intuition.

One option here would be to try to extrapolate the value of the potential productivity measure with respect to larger corpus sizes, using statistical models (LNRE models, for ‘Large Number of Rare Events’ models) that are appropriate for the Zipfian properties of word frequency distributions, as done in Baayen (2001). However, Evert and Lüdeling (2001) provide a detailed analysis of the specific problems encountered in automatic pre-processing of large amounts of textual data, and conclude that in the general case automatic tools deliver results that are too noisy to yield reliable extrapolations of potential productivity measures.

For this reason we choose instead to focus on the two existing corpora and explore how potential productivity measures vary when constraints are imposed on the minimum number of tokens that instantiate the morphological category whose potential productivity we want to measure.

We compute potential productivity measures for the full Wikipedia corpus by setting a minimum threshold for the number of tokens that instantiate morphological categories, starting from 0 (no constraint) and proceeding in increments of 5 up to 100. For each threshold level we derive productivity rankings.

We consider first a variant on the notion of stability of the productivity ranking by computing Spearman’s rank correlation between the ranking obtained when setting the minimum threshold value to 100 and each of the rankings obtained at the 5-increment interval, shown in Figure 5.

Figure 5 visualizes how frequency effects impact the stability of potential productivity rankings.

When the minimum threshold is very low (0 or the first few increments above 0), productivity rankings are extremely unstable, to the point that the correlation is negative with respect to the final ranking based on the highest threshold value (100). This is due to the fact that at low thresholds many categories have productivity 1, since all the types for those categories are instantiated by exactly 1 token. Because no attempt is made to add a second sort order (which could be for example the number of types) to the productivity value, the top portion of the ranking is basically random, since there can be hundreds of categories with productivity 1. However, as the constraint on the minimum threshold becomes increasingly stronger, rankings tend to stabilize. For the last few thresholds values the rankings at the highest positions are very similar. For the threshold value 100 the ranking is shown in Table 3.

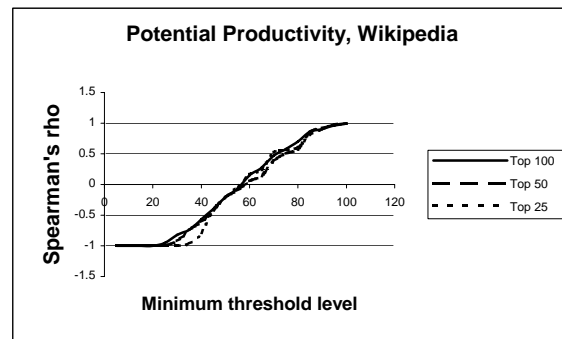


Figure 5: Potential productivity stability as a function of threshold level, on the Wikipedia corpus.

A simple visual comparison between Table 3 and Table 2 shows that complex adjectives with the highest potential productivity are very different from those with the highest realized and expanding productivity.

The main qualitative impression is that adjectives with high potential productivity seem to exemplify patterns that are grammatically and compositionally very transparent. There are many categories based on adverbs (‘partially-X’) and past participles (‘X-obsessed’). These patterns could indeed be written out as two separate words (with a space instead of the hyphen) forming a syntactic adjective phrase. On the other hand, adjectives with high realized and expanding productivity seem to exemplify patterns that tend towards lexicalization. Patterns like ‘anti-X’, ‘pre-X’, ‘post-X’, ‘re-X’, ‘semi-X’, ‘multi-X’ could be indeed written

out as a single word that is syntactically an adjective⁴. In other words, we could suggest that potential productivity might correlate with the early stages of a lexicalization process, and realized and expanding productivity might correlate with the more advanced ones. In this paper we will not elaborate this suggestion any further, and leave it for future work.

Wikipedia potential productivity rankings			
1.X-ish	0.735849	11.X-wearing	0.47619
2.almost-X	0.649123	12.X-penned	0.451327
3.easily-X	0.613208	13.un-X	0.451001
4.already-X	0.581522	14.X-centric	0.443662
5.X-obsessed	0.529412	15.nearly-X	0.442308
6.partially-X	0.526786	16.X-kilometer	0.441176
7.X-focused	0.508197	17.X-associated	0.437229
8.X-less	0.48913	18.X-capable	0.434066
9.X-inspired	0.48537	19.previously-X	0.427907
10.micro-X	0.481081	20.mini-X	0.42449

Table 3: Potential productivity rankings for the Wikipedia corpus, based the threshold value 100.

We consider now another angle of the relationship between potential productivity on one hand and realized and expanding productivity on the other. Using the same rankings obtained by varying minimum threshold for the number of tokens that instantiate morphological categories, we calculate the Spearman’s rank correlation of each potential productivity ranking at every increment interval with respect to correspondent rankings for realized and expanding productivity on the full Wikipedia corpus.

Figure 6 shows that no matter how we set the minimum threshold level, rankings for potential productivity appear to be significantly different with respect to the rankings for realized and expanding productivity, and at least for the size of the Wikipedia corpus the correlation remains strongly negative.

A slight upward trend is detectable though, and this corresponds to the fact that setting the minimum threshold value to 100 increases somewhat the strength of the correlation between potential productivity and the other two types as corpus size increases (as we discussed in regard to Figure 4).

⁴ The occurrence of single orthographic word variant can be construed as a signal of lexicalization for complex adjectives, but not in the general case, as demonstrated by the the orthographic integrity of words within idioms.

Indeed, the values of ρ at the rightmost edge of the respective curves in Figures 4 and 6 are the same, since they corresponds to the rank correlation among productivity rankings for the full Wikipedia corpus with minimum threshold value set to 100. We can interpret this slight upward trend as a consequence of the mitigation of frequency effects (which typically affect potential productivity rankings, as discussed above) brought about by the raising of the minimum threshold level.

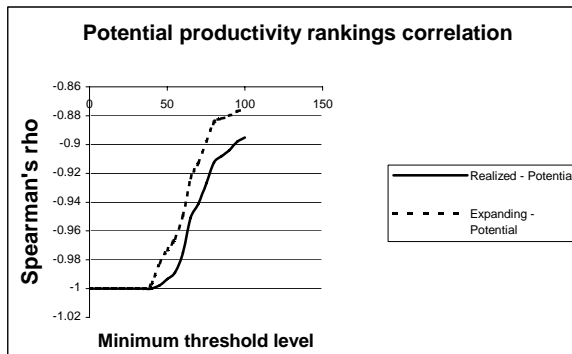


Figure 6. Potential productivity rankings correlation as a function of minimum threshold level in Wikipedia

This behavior highlights the fact that, compared to the other two measures, potential productivity focuses by design on a significantly different facet of the notion of morphological productivity. Potential productivity attempts to capture in a simple formula the ability of a morphological category to *continue* to enable speakers to generate new words. It is intended to characterize a morphological category in terms of the number of potential words that can still be created in that category. Thus it could be the case that a morphological category is well established in the language (realized productivity) and is expanding (expanding productivity), but does not have much potential for further expansion (potential productivity).

In order to explore these scenarios, we compare the three measures for the specific morphological pattern ‘non-X’, which has both the highest realized and expanding productivity for all the 19 intermediate corpus slices, and for the full corpus size (both for the Wikipedia corpus and for the Web corpus). On the Wikipedia full corpus this category yields 6444 distinct types for a total of 47489 tokens (including 4496 *hapax legomena*), ranging from very frequent types such as *non-profit* (2268 tokens), *non-existent* (619 tokens),

non-standard (501 tokens) to *hapax legomena* such as *non-ratified* (1 token), *non-amphibious* (1 token), etc. Specifically, the fact that ‘non-X’ has the highest expanding productivity means that it has more *hapax legomena* than any other morphological pattern. This notwithstanding, the fact that its different types yield a very large number of tokens results in a very low potential productivity value of ‘non-X’ (on the full corpus). Setting the minimum threshold value at 50, the average potential productivity rank for ‘non-X’ steadily gets worse as corpus size increases from rank 24 for the first intermediate corpus to rank 2404 for the full size corpus, in stark contrast to the rank position 1 it has for realized and expanding productivity (we use ranks since absolute productivity values for different measures cannot be compared to each other).

On the other hand, the pattern ‘X-ish’ which has the highest potential productivity when the minimum threshold is set to 100 exhibits radically different behavior. On the Wikipedia corpus this category yields 99 distinct types for a total of only 112 tokens, but including 87 *hapax legomena*. The most productive types are *cartoon-ish* and *blue-ish* (with 3 tokens each) and *40-ish*, *blues-ish*, *vanity-ish*, *place-ish*, *bully-ish*, *punk-ish*, *tree-ish*, *white-ish*, *noir-ish* (with 2 tokens each). All other types have 1 token each. The high proportion of *hapax legomena* pushes ‘X-ish’ at the very top of the potential productivity ranking. As is expected, the realized and expanding productivity ranks are much lower. However, they do increase steadily from rank 5204 and 3990 respectively for the first intermediate corpus to rank 403 and 242 for the full size corpus. We can interpret this trend as a consequence of the fact that for the category ‘X-ish’ almost every new token is also both a new type and a new *hapax legomena*. The effect of this is that it slows down the rate of decrease of the realized and expanding productivity measures (which always decrease inversely to corpus size), thus lifting the rank of ‘X-ish’ above that of competitors as corpus size increases.

Note that from a linguistic point of view, the hyphen in potential productivity patterns seems to signal the awareness of the writer/speaker regarding the ‘novelty’ of the complex adjective, along a continuum from compounding (‘almost-X’, ‘X-capable’) to more established derivational morphemes (‘X-ish’, ‘un-X’).

5 Conclusions and future work.

A general trend has consistently emerged throughout the analysis, as illustrated by the opposite properties of the categories ‘non-X’ and ‘X-ish’. We have shown from a variety of angles how realized and expanding productivity measures tend to capture closely related aspects of morphological productivity, both in terms of the specific words and global correlation between productivity rankings. On the other hand, potential productivity consistently exhibits clearly different and indeed opposite properties compared to the other two, again both in terms of the specific words that score the highest values and in terms of productivity ranking correlation patterns.

The most significant question to be addressed next is the extent to which the above-mentioned conclusions about the relationships among different types of morphological productivity carry over to other segments of language.

Regarding complex adjectives, future work will focus in more detail on their internal typology, specifically considering the extent to which they impose selectional restrictions (whether morpho-syntactic or semantic in nature) on the component associated with the base generator of a category, and the correlation patterns between such constraints and the three types of morphological productivity.

References

- R. Harald Baayen. 1993. On Frequency, Transparency, and Productivity. In: Geert E. Booij and Jaap van Marle (eds). *Yearbook of Morphology*. Dordrecht: Kluwer Academic Publishers, 181-208.
- R. Harald Baayen. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer Academic Publishers.
- R. Harald Baayen. 2006. Corpus Linguistics in Morphology: Morphological Productivity. In: Lüdeling, A. and M. Kytö (eds). *Corpus Linguistics. An International Handbook*. New York: Mouton De Gruyter.
- Stefan Evert and Anke Lüdeling. 2001. Measuring Morphological Productivity: Is Automatic Preprocessing Sufficient? In: *Proceedings of the Corpus Linguistics 2001 Conference*, 167-175.
- Karo Moilanen and Stephen Pulman. 2008. The Good, the Bad, and the Unknown: Morphosyllabic Sentiment Tagging of Unseen Words. In: *Proceedings of ACL 2008*.