

A Corpus Based Study on Productivity Measurement of Suffixes ‘-ous’ and ‘-ish’

Mengxuan Bai^{1,a,*}

¹English Department, University College London, WC1E 6BT, United Kingdom
a. uclebai@ucl.ac.uk

*corresponding author

Abstract: In recent years, a growing number of studies have been conducted on the productivity of word formation. ‘-ous’ and ‘-ish’ are common adjective suffixes which have the same general meaning of ‘to be similar to something’ and are consequently regarded as synonymous affixes. However, most prior studies have addressed lexical productivity from a single perspective, such as diachronic analysis or synchronic measurement, while few have combined the above aspects to conduct multi-angle studies. The comprehensive research perspective adopted in this paper provides several novel ideas for developing a further understanding of the productivity of the two affixes, and even for the wider study of morphological productivity.

Keywords: Productivity, ‘-ous’, ‘-ish’, Multi-angle, Word formation

1. Introduction

The relevance of productivity in the study of language and communication is obvious. Since then, research on productivity in morphological structures has flourished, and it remains one of the field's most hotly debated topics. In order to help students of a second language acquire new vocabulary and create their own terms, an understanding of the rules of creation and production in the language is crucial [1]. Methods for measuring synchronic productivity will be used as a starting point for this study. Notable factors like as the hapax ratio, token frequency, and token kind will also be taken into account. Second, this paper uses a diachronic perspective to look at how the production of the two suffixes has changed throughout the years. Finally, this thesis will summarise the potential limitations of estimating the productivity of morphological processes and conduct a simultaneous analysis of the various results from both synchronic and diachronic perspectives.

2. Literature Review

2.1. Definition of Productivity

The concept of productivity is fiercely debated within the field of morphology. Productivity may be thought of in a broad sense as the amount of imagination that can be put into making up new words and phrases through a system of rules [2]. Morphological creation, or the making of new words, has been the focus of the vast majority of studies.

This research relies on Plag's idea of productivity as its theoretical basis because of his thorough examination of several productivity definitions. According to Plag, the term "productivity" is used to characterise the quality of an affix or morphological process that consistently generates new words in morphological texts [3]. Examples include Schultink, who labels production as a morphological phenomenon [4]. Morphological productivity, as he describes it, is the seemingly boundless ability of language users to generate new words in their heads by using existing words' forms and meanings in lexical processes. Also, according to Bauer, a lexical process's productivity is measured by its ability to consistently produce novel words [5]. It's worth noting that Bauer divides production into qualitative "availability" and quantitative "profitability." [6]. The term "availability" refers to the possibility of using a certain lexical process to generate new words in a specific language at this moment. For instance, in English, the prefix 'ise' is considered accessible since it may be used to create new words, but the prefix 'en' is no longer used and is thus unavailable. Therefore, availability is a binary property. On the other hand, a lexical process's profitability is proportional to the number of novels, and suitable word forms it produces [7]. In addition to eliminating intentionally created terminology, Bauer also eliminates abbreviated words, blends, and initials from productivity [8]. Therefore, only morphological processes that adhere to regulations are considered productive. The question of whether or not the creation of words is deliberate or constrained by rules is a contentious one in the academic community.

2.2. Original Meaning

Popular adjective endings include both '-ous' and '-ish'. The Oxford Online English Dictionary (2019) explains that the suffix '-ous' comes from the Latin '-sus' and that it was heavily influenced by the ancient French '-ous' and '-eux', which mean "full of" or "having the essence or character of" (e.g. fame-famous, delice-delicious). It may also mean molecules with a lower valence in chemistry than those ending in "-ic" (2019). The suffixes '-eous,' '-ious,' and '-uous' may also be formed by combining it with the prefixes 'e-', 'i-', and 'u-' (e.g., vitrum - vitreous, trepidate - trepidatious, ambiguous).

The suffix '-ish' is derived from a number of Germanic languages, including Gothic (where it appears as '-isks'), Old Norse (where it appears as '-iskr'), Old High German, Old Saxon, Old Frisian, and Old English, etc. In Old English and related languages, the -le suffix is most often added to nouns that signify "belong to someone or something" or "have the essence and character of someone or something," such as idiot and stupid. Today, the suffix is commonly added to the end of a noun or a noun phrase to create an adjective that describes the noun. Similar to or possessing the qualities of; examples include thinnish (slightly thin), coldish (slightly chilly), and dullish (not very interesting) as adjectives (slightly blunt). As an adjunct to an hour or year, it denotes "approximately," as in "elevenish."

It is clear that the adjectival suffixes '-ous' and '-ish' have been around for quite some time. As a result, it is appropriate to use a diachronic research perspective in this thesis.

3. Methodology

With a combined capacity of one billion words and a wide variety of text genres, including spoken, fiction, magazine, news, and academic, extracting a corpus from this work is a breeze. The Oxford English Dictionary (OED) is also utilised for corpus retrieval and statistics. This research manually screened 1000 words based on the number of ous- and -ish-derived word morphs due to the database's limited filtering capabilities and the large number of words collected.

Approximately 2,000 words were selected for further examination based on the number of morphemes they contained that ended in ous or -ish.

Analysis of the temporal productivity of -ous and -ish constructions is greatly aided by the fact that the Oxford Online Dictionary of English, which contains roughly 800,000 words since the early 2th century, clearly indicates when and how many words formed by -ous and -ish were added for the first time, as well as the number of words added in each century. In addition, a chi-square test of the Hapax ratios was performed using the statistical software SPSS 23 to see whether there is a statistically significant difference in the word creation productivity of the two suffixes.

4. Results and Discussion

The creation of the English adjectival suffixes '-ous' and '-ish' are compared and contrasted, both now and historically. In addition to discussing the difficulties of assessing morphological productivity, this chapter also addresses some of the issues that have been raised in the literature.

4.1. Synchronic Measurement

This study uses transparency-based measures of productivity, hapax-based measures of productivity, and relative frequency-based measures of production to examine synchronicity in the workplace.

4.1.1. Method Based Transparency

The term "transparency" is used in this article to describe how little the pronunciation and meaning of the affix and related lexeme change throughout construction. COCA found that the following 10 words ending in '-ous' or '-ish' occur most often in their database.

Table1: Word endings in -ish and -ous are often used to form new words. (source: COCA).

ish		ous	
Span	Spanish	fame	famous
ten	tenish	curiosity	curious
Jew	Jewish	mystery	mysterious
Poland	Polish	continue	continuous
style	stylish	anxiety	anxious
outland	outlandish	caution	cautious
green	greenish	rigor	rigorous
boy	boyish	infection	infectious
blue	bluish	vigor	vigorous
book	bookish	marvel	marvelous

Table 1 shows that the new words' meanings are a combination of those of the suffixes and the word bases, suggesting that both affixes display semantic and phonemic transparency. Harmful, dangerous, magnificent, brilliant, stupid, and youngish are all pairs that may be used interchangeably.

Plag suggests that transparency increases output [3]. The main idea is that the words used in entailment techniques are often easy to understand both semantically and phonologically. In other words, it is possible to infer the meaning of a semantically transparent word by examining its morphemes. Low levels of phonological integration between affixes and stems characterise the phonological structure of transparent word forms. In contrast to the semantic drift and opacity that are typical in the access and storage situations of whole words, productive categories generally leave little leeway for these kinds of ambiguities since their derivatives favour a deconstructive approach [9]. The deconstruction method is useful because it improves the semantic representations of the individual morphemes. More proof that efficiency and phonetic precision go hand in hand may be

found in Baayen's work [10]. Baayen finds a strong correlation between junctural phonotactics and affix production based on an analysis of the phonemic boundary phonological structure (phonotactics) of eighty affixes [10]. This study challenges the conventional wisdom that ordinary suffix formation does not include phonemic transitions. Therefore, the deconstruction ability and output level may be affected by the phonemic structure of an affix (and the structure of the allophone it generates).

In conclusion, the efficiency of developing new words is proportional to how transparent they are, i.e. the more transparent they are, the more efficient the process is. In this respect, both prefixes might be seen as quite useful.

4.1.2. Method Based on Hapax Ratio

Baayen recommends hapax legomena as a means of evaluating the efficacy of current construction practises [11]. This research proposes that hapaxes are the most fertile ground for the development of new words. In other words, the proportion of an affix among the hapaxes in a corpus is an indicator of strong production, and the number of hapaxes and neologisms within a specific morphological category are correlated [12]. Since the number of neologisms in a given morphological category is equal to the number of hapaxes in that category, the number of hapaxes may be thought of as a measure of output, as stated by Plag et al. [13].

Baayen recommends using the ratio of hapaxes with a certain affix to the total number of words with that affix as a measure of productivity [11]. Productivity (P) = Number of Hapaxes (n) / Total Number of Words (N) where n is the number of hapaxes for the affix in question and N is the total number of words that might include that affix. According to Baayern, an affix can be understood as follows: an affix with a large number of hapaxes has a large P-value and high productivity, whereas an affix with a large number of high token-frequency words has a small P-value and poor production [11]. This fits with a commonsense understanding of production [14]. However, type frequency is not taken into account by this technique. Therefore, this study uses a control variable approach to examine a corpus of 1,000 '-ous' and '-ish' derivatives.

For the 2,000 words that were recovered and manually filtered using COCA, this research indicates that the total number of tokens derived from '-ous' is 152354 (N), and the number of hapaxes is 121. (n). There are 429 hapaxes and 123686 tokens that end in '-ish' (n). The results of both prefixes are shown in Table 2 and Figure 1.

Table2: Statistics for 2000 words derived from '-ous' and '-ish'.

	Type	Token (N)	Hapax (n)	P = n/N%
ous	1000	152354	121	0.079%
ish	1000	123686	429	0.347%

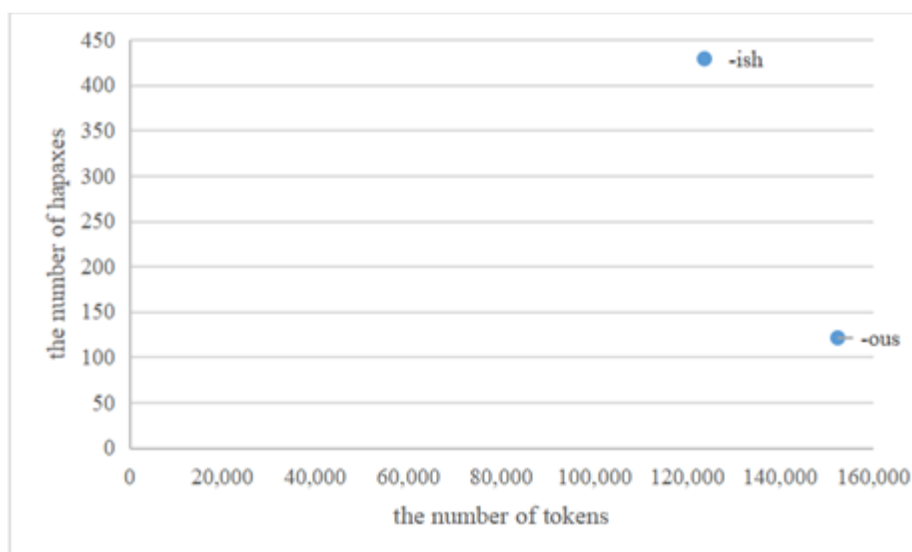


Figure1: Distribution of the number of hapaxes and tokens of words derived from ‘-ous’ and ‘-ish’.

Token counts show that for a set of word classes, '-ous' derivatives are much more common than '-ish' derivatives. However, whereas '-ous' affixes only had a hapax ratio (P-value) of 0.079%, '-ish' affixes had a hapax ratio (P-value) of 0.347%, almost four times greater. This study applies a chi-square test to check whether the hapax ratio of the two suffixes is independent, further supporting the projected findings. The chi-square test shows that the hapax ratios of words formed with the two affixes are quite different ($p < 0.05$). There will be fewer morphemes in the corpus and a higher hapax ratio if more new words are generated and the rate of morpheme creation is high. The morphological output of '-ish' is far greater than that of '-ous,' measured in terms of hapax ratios.

4.1.3. Method Based on Relative Frequency

Accessing the individual morphemes (the deconstruction method) or simply accessing the whole-word representation are the two approaches to store and retrieve morphologically complicated words, as stated by McQueen and Cutler [15]. (the whole-word route) Some researchers think these two channels work together and compete with one other, while others believe only one route may be used at once. You may learn more about these paths in Chapter 2.

The most crucial factor in determining which phrase is used for which route is how often it is used [16]. Resolution for a given target word is based squarely on the frequency ratio between the derivative and the word basis. According to the second chapter, relative frequency is calculated by multiplying the derivative frequency by the base frequency.

This work makes a number of adjustments to the standard procedure in order to account for the fact that its emphasis is on affix productivity rather than productivity at the word base. In this study, we use the frequency of word base types as our major reference variable. Affixes may attach to a wide variety of word bases, which is represented by this variable [17]. If there are fewer viable word stems from which to build new words, then word creation productivity will be lower.

Table 1 shows that the vast majority of 'ous'-capable terms are non-specific concepts like "fame," "danger," "ambition," etc. Unlike abstract nouns like "English" or "elevenish," "yellowish," "bullish," or even "Johnish," most nouns to which "-ish" may be appended are concrete nouns like names, numbers, colours, and animals. It's also possible to add -ish to the end of an adjective or even a whole phrase, as in youngish (young + ish). This study can only suggest that 'ish' affixes may be connected to more lexical categories than 'ous' affixes because of the paucity of research comparing the lexical

quantities of abstract and concrete nouns. However, it is difficult to determine whether or not the two construction processes significantly differ in the frequency of word base types. productivity.

4.2. Diachronic Measurement

Fewer research have looked at how our vocabulary has changed throughout time. Thus, the purpose of this study is to examine the historical development of neologisms (new lexical categories) from their inception through their eventual demise.

4.2.1. Diachronic Change of Productivity

To determine how the rate at which new words are being added to the dictionary has changed over time, I counted how many new words were added to the dictionary in each century and what percentage of those new terms were the result of each lexical process. The results of this evaluation are shown in Figure 2.

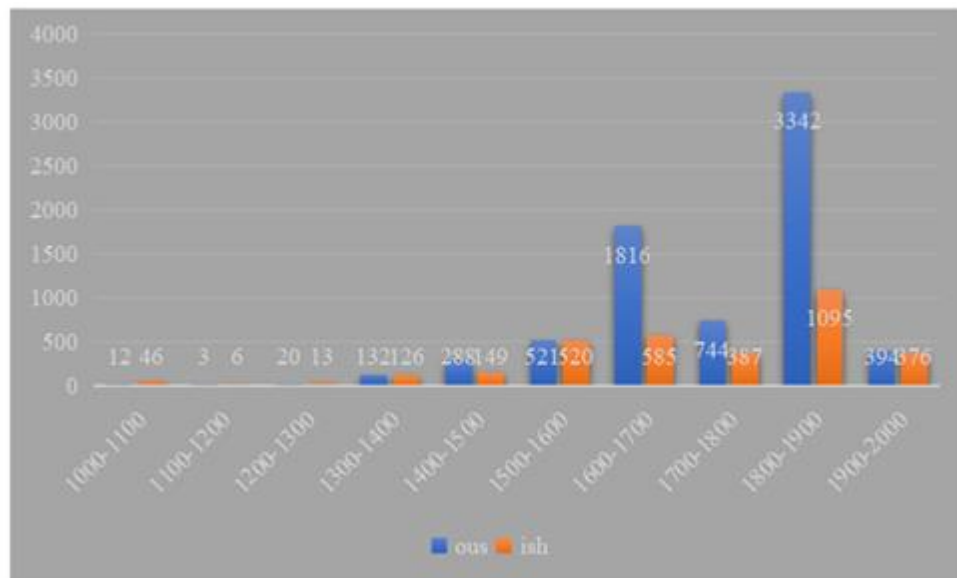


Figure2: Distribution of the number of words derived from -ous and -ish for initial inclusion in the OED by century, 1000-2000.

Figure 2 shows that the number of derived words with the suffixes "-ous" and "-ish" is generally on the rise and falling, respectively. The number of derived words first stated for both is lowest in the 12th century, then quickly climbs in the 17th, peaks in the 19th, and then significantly drops starting in the early 20th century. Words ending in '-ous' also outnumbered those ending in '-ish' by a wide margin throughout the 17th and 19th centuries. Below is a graph showing the percentage of words from each century between 1000 and 2000 that end in the suffixes "-ous" and "-ish."

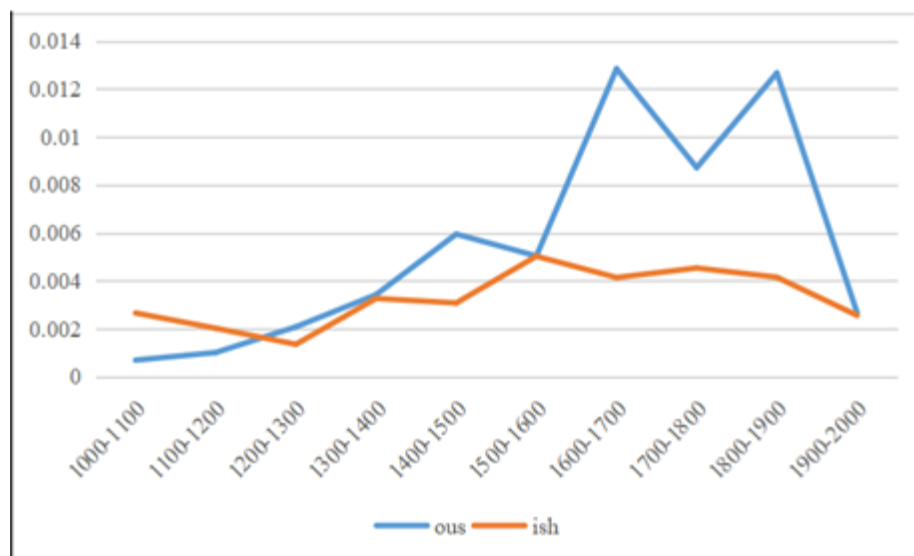


Figure3: The number of derivatives from ‘-ous’ and ‘-ish’ as a proportion of the total number of words recorded each century, 1000-2000.

The calculations reported at the beginning of this chapter are used in Figure 3 to illustrate the temporal patterns in the productivity of the two adjectival suffixes. As a rule, the '-ish' process has a lower rate of productivity, with the maximum rate happening between the 16th and 17th centuries at 0.5% and the lowest rate occurring between the 13th and 14th centuries at 0.14%. The productivity of the '-ous' building process, on the other hand, fluctuates widely, reaching a high of 1.28 percent in the 17th and 18th centuries and a low of 0.10 percent in the 12th and 13th. The results of this calculation, in contrast to those obtained using the aforementioned synchronic method, are used only to account for variations in the productivity of a given lexeme as it has evolved over time.

This thesis presents the concept of utility in order to conduct a more thorough analysis of the aforementioned diagram. The degree to which a newly acquired word may be used to construct a new term when needed by the speaker is at the heart of this idea. The efficiency of an affix is related to how useful it is. Figures 2 and 3 show that the efficiency of the two affix production procedures changes on a regular basis, which suggests that the usefulness of a process varies with the need for the creation of new words using it. The reasons for this shift in demand are not discussed since doing so would go beyond the subject of this research. Findings from this research show that both of these processes are very useful and productive, as seen by the large number of words generated by the addition of these affixes. Also, the results show that the demand for the two processes in this study has remained relatively constant, despite the fact that '-ous' changes more than '-ish,' with an increase in the number of first-introduction words in the seventeenth and nineteenth centuries, followed by a decline in the demand for both constructions and, thus, their usefulness in the eighteenth and twentieth centuries. Since the 13th century, the graph shows that the '-ish' process has been less useful than the '-ous' process.

5. Conclusion

Following are four conclusions drawn from a comparison of the effectiveness of the '-ous' and '-ish' constructions: 1) Productivity of the former is higher during periods of high demand, such as the 17th and 18th centuries, while that of the latter is higher during periods of lower demand, such as the 13th century and the 16th and 17th centuries. Both constructions' productivity has been on the decline since the 20th century. There was a nadir for both processes before the 13th century, and they've dropped down significantly since the 20th century. Two) Neither method compromises the integrity

of the morpheme's original meaning or phonetic shape, yet both are very productive. 4) The latter may be linked to a wide range of base types in terms of frequency, but it's hard to establish a direct comparison between the two processes due to a lack of specific data on concrete and abstract nouns. Fourth, the latter is far more productive than the former in terms of Hapax ratios ($p < 0.05$). The morphological output of the two adjectival suffixes is compared and contrasted quantitatively in this research.

The purpose of this study is to benefit second-language learners of vocabulary via corpus and word-creation procedures by analysing and statistically comparing the word formation productivity of the two adjectival suffixes in a number of ways. However, a quantitative comparison of the efficiency of the two word production processes is particularly challenging due to the absence of specific data on the essential lexical categories, therefore only the kinds of attachable words are compared here. In this analysis, the Hapax ratio served as the single quantitative metric for drawing statistically significant differences between the two word creation procedures. Due to the large amount of information in the COCA corpus, which would have to be manually filtered out, we only have a tiny sample size to work with. The hope is that studies will grow and supplement one another.

References

- [1] Sproat, R., & Shih, C. (1996). A corpus-based analysis of Mandarin nominal root compound. *Journal of East Asian Linguistics*, 5(1), 49-71.
- [2] Anderson, S. R. (1982). Where's morphology?. *Linguistic inquiry*, 13(4), 571-612.
- [3] Plag, I. (2006). *Word-formation in English*. Cambridge University Press.
- [4] Schultink, H. (1961). Produktiviteit als morfologisch fenomeen. In *Forum der letteren* (Vol. 2, No. 1, pp. 110-125).
- [5] Bauer, L. (2005). Productivity: theories. In *Handbook of word-formation* (pp. 315-334). Springer, Dordrecht.
- [6] Bauer, L. (2001). *Morphological productivity* (Vol. 95). Cambridge University Press.
- [7] Aronoff, M. (1976). *Word formation in generative grammar*. *Linguistic Inquiry Monographs* Cambridge, Mass, (1), 1-134.
- [8] Bauer, L., & Laurie, B. (2002). *English Word-Formation*. Cambridge university press.
- [9] Baayen, H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study.
- [10] Baayen, R. H. (2009). 43. Corpus linguistics in morphology: morphological productivity. *Corpus linguistics. An international handbook*, 900-919.
- [11] Baayen, H. (1992). Quantitative aspects of morphological productivity. In *Yearbook of morphology 1991* (pp. 109-149). Springer, Dordrecht.
- [12] Nishimoto, E. (2003, February). Measuring and comparing the productivity of Mandarin Chinese suffixes. In *International Journal of Computational Linguistics & Chinese Language Processing*, Volume 8, Number 1, February 2003: Special Issue on Word Formation and Chinese Language Processing (pp. 49-76).
- [13] Plag, I., Dalton-Puffer, C., & Baayen, H. (1999). Morphological productivity across speech and writing. *English Language & Linguistics*, 3(2), 209-228.
- [14] Bakken K. (1998). *Leksikalisering av Sammensetninger*. Oslo:Universitetsforlaget.
- [15] McQueen, J. M., & Cutler, A. (1998). Spotting (different kinds of) words in (different kinds of) context. In *The Fifth International Conference on Spoken Language Processing* (pp. 2791-2794). ICSLP.
- [16] Hay, J. (2001). Lexical frequency in morphology: Is everything relative?
- [17] Plag, I. (2012). Morphological productivity. In *Morphological Productivity*. De Gruyter Mouton.