Narrow Productivity, Competition, and Blocking in Word Formation

Junya Morita
Kinjo Gakuin University
College of Humanities
morita@kinjo-u.ac.jp

Abstract

The present study explores the productivity of word formation processes in English, focusing on word composition by suffixes such as *-ize* (e.g. *transcendentalize*), *-(a)(t)ion* (*territorization*), and *-al* (*realizational*). An optimal productivity measure for affixation is identified, which makes best use of hapax legomena in a large-scale corpus and attaches great importance to the base forms of an affix. This measure is then applied to the data collected from a large corpus to compute the productivity values of twelve kinds of affixes. The detailed investigation reveals that (i) the high productivity rate of an affix demonstrates a creative aspect of the affix, giving full support to the idea of "generative" morphology, (ii) productivity is gradient; very high, fairly high, and low productivity of affixes are recognizable, and (iii) this is necessarily reflected in determining the word form of a derivative (cf. *territoriz<u>ation</u>*); competition is carried out to decide which affix is selected for a given base form (*territorize*) and the "losers" (*-ment*/*-al*) are blocked out.

1. Introduction

Productivity – the potentiality of creating new lexemes – is one of the central themes of morphological studies and attempts have been made to establish a refined and accurate productivity measure (Baayen and Renouf, 1996; Plag, 1999). The productive devices of word formation create new items whenever the necessity arises, and the regular selection of a productive affix from the competing affixes enables us to describe the lexicon elegantly. The aim of this study is to measure the productivity of word formation patterns based on a large parsed corpus of natural language and show its theoretical implications. The outline of this study is as follows: after pointing out some previous approaches to productivity measurement and their problems, we propose a new measure in §2. Section 3 applies this measure to the data obtained from a large-scale corpus to calculate the productivity of twelve kinds of suffixations. Section 4 concentrates on the results of the research and their theoretical implications.

2. Productivity Measure

2.1. Previous Studies and Their Problems

Speakers always have the capacity to make up new words and word formation devices are crucially engaged in generating new forms (Aronoff, 1976). It is important to recognize what sort of devices speakers have at their command and to what extent each device can produce new items. It is thus necessary to identify the productivity of each device; the extent to which a word formation device can give rise to new words (Lieber, 2010: 59).

There have been three major approaches to quantifying productivity. The first way is to simply count the total number of the relevant complex words listed in a large dictionary; *The Oxford English Dictionary* contains more *in-* negatives than *non-* negatives, and so the former prefixation is judged to be more productive than the latter. This approach has the problem of neglecting the fact that transparent derivatives are unlikely to be listed in dictionaries (Plag, 1999: 98). The second way of quantifying productivity is to count the number of the relevant complex words which are used in a given material. This way also has the drawback of taking no notice of the fact that there exists an affix which no longer yields new words although the related words are still used; the nominal suffix *-th* has

Keywords: affixation, productivity, hapax, generative morphology

no capability of making new words even though a set of nouns in *-th* still exist (Plag, 1999: 22-23).

The last productivity measure attaches great importance to hapax legomena – token frequency 1 – of a large-scale corpus (Baayen and Renouf, 1996; Plag, 1999). This is based on the view that the capacity of an affix to create new forms crucially involves the degree to which the affix produces words of very low frequency (Hay, 2003). Baayen and Renouf (1996: 73) propose a productivity measure: *Productivity* $(P)=n_1/N$, where n_1 is the number of hapaxes and N is the total number of tokens. For instance, as seen in Table 1, when the token number of *-ize* is 20865 and the hapax number of *-ize* is 80, the productivity value of *-ize* is 0.0038.

suffix	hapaxes (n_1)	tokens (N)	types (V)	productivity (P)
-ate	69	41561	481	0.0017
-ify	18	7236	88	0.0025
-ize	80	20865	347	0.0038

Table 1: The productivity of *-ate*, *-ify*, and *-ize* in the Cobuild corpus. (cf. Plag, 1999: 111)

This corpus-based model also has weaknesses: (i) the existence of extremely highly frequent derivatives (e.g. *realize*, 5506 tokens in the British National Corpus (BNC)) excessively lowers the productivity value of an affix, preventing us from finding the real value of productivity, and (ii) to calculate the productivity value of the whole affixation process would be of little significance to morphological theory, as will be discussed shortly.

2.2. A Proposal

We basically accept the corpus-based productivity measure proposed by Baayen and Renouf, 1996. In adopting it, however, we will revise it in two respects. First, for the productivity formula given above, the total number of types (but not tokens) is placed in the denominator. This is derived from the view that the productivity of a particular process is reflected in the type frequency of the process (Goldberg, 1995: 134-139). Given the observation of *-ize* suffixation in Table 1, the revised productivity formula "$P=n_1/V$ (V: types)" is applied, giving a productivity value of 0.2305. In this new measure, the productivity of *-ize* is defined as the potentiality of creating 80 kinds of new words when 347 kinds of *-ize* derivatives are used. This is in sharp contrast to Baayen and Renouf's framework, where the productivity of *-ize* is defined as the potentiality of coining 80 kinds of new words when *-ize* words are used 20865 times.[1]

The second point is that productivity should be measured in a small and specific domain. An affix has a tendency to combine with certain affixes (Jespersen, 1949: 449). Based on this fact, Aronoff (1976: 36) claims that "… there is no absolute way to say that one word formation rule is more productive than another. Rather, one must take into account the morphology of the base." Valuing this base-motivated derivation, we propose the "narrow productivity" of an affix: the rate of its deriving new words for a given base form. For example, of a total of 106 types in *[[X-al]-ize]* collected, 42 *-ize* hapaxes are detected in BNC, giving a productivity value of 0.396 for the *X-al* base. In the next section, we will morphologically classify the bases of each affixation and calculate the productivity value of affixation for each base.

3. A Research

3.1. Target and Methodology

We calculate the productivity values of (i) five nominal suffixes (*-(a)(t)ion/-al/-ment/ -ity/-ness*), (ii) four adjectival suffixes (*-al/-less/-ic/-ical*), and (iii) three verbal suffixes (*-ize/-ify/-ate*) for the total number of twenty kinds of base forms (e.g. *X-ize/X-ify*). In order to collect the hapaxes ending in these suffixes, we look for them in BNC, a 100-million-word corpus. By repeatedly using the "wild card" function of a research engine, the frequency of complex words ending in the above suffixes is checked

[1] Baayen and Lieber (1991: 810-811) refer to several possible methods to gauge productivity, one of which is essentially equivalent to our productivity measure.

to find the ones which occur only once in the corpus.[2] As for ascertaining the total number of types of the derivatives concerned, we make a list of those which are included in Lehnert's *Reverse Dictionary of Present-Day English* and attested in BNC.[3]

3.2. Result

The results of the research are put into tabular forms in terms of (i) noun-forming suffixes, (ii) adjective-forming suffixes, (iii) verb-forming suffixes, and (iv) a suffix which is sensitive to a particular prefix.

3.2.1. Noun-forming Suffixes

Table 2 shows the total number of types, the total number of hapaxes, and the productivity values concerning the deverbal nominal suffixes *-(a)(t)ion*, *-ment*, and *-al* for the bases of *X-ize*, *X-ify*, and *X-ate*. The number of types and that of hapaxes of *[[X-ize]-(a)(t)ion]* nouns are 252 and 47 respectively, giving its productivity value of 0.187. By contrast, since *[[X-ize]-ment]* or *[[X-ize]-al]* nouns are not found in BNC, the suffixes *-ment* and *-al* take a productivity value 0 for these base forms.

	1. -(a)(t)ion			2. -ment/ 3. -al		
base	V	n_1	P	V	n_1	P
X-ize	252	47	0.187	0	0	0
X-ify	84	31	0.369	0	0	0
X-ate	460	30	0.065	0	0	0
X (simple)	391	7	0.018			

e.g. 1. *Mongolization*, *humidification*, *fantastication*, *spoilation*

Table 2: The productivity of deverbal-noun forming suffixes.

Table 3 displays the type number, hapax number, and productivity values concerning the deadjectival nominal suffixes *-ity* and *-ness* for the bases of *X-able*, *X-al*, *X-ic*, *X-ile*, *X-ar*, *X-ous*, and *X-ive*. We find that the productivity of *-ness* is slightly lower than that of *-ity* for the base *X-al*, although overall *[[X-al]-ity]* types are much more than overall *[[X-al]-ness]* types.

	4.-ity			5.-ness		
base	V	n_1	P	V	n_1	P
X-able	386	177	0.459	19	4	0.211
X-al	118	52	0.441	17	5	0.294
X-ic	35	19	0.543	2	1	0.500
X-ile	26	6	0.231	1	1	1.000
X-ar	24	17	0.708	0	0	0
X-ous	69	10	0.145	170	58	0.341
X-ive	29	7	0.241	88	26	0.295

[2] For this hapax-finding I am indebted to the research engine of http://view.byu.edu/reg3. asp?c=aybfiyfml.

[3] For the purpose of deciding which suffix is selected in creating a new word, we redefine a hapax as a word which occurs only once and whose rival formation is of zero frequency. Thus, when the frequency of *etymologic* is 1 and that of *etymological* is 29, *etymologic* is not regarded as hapax.

X (simple)	33	3	0.091	321	29	0.090

e.g. 4. *allowability, ornamentality, ellipticity,* 5. *mathematicalness, patheticness, rifeness*

Table 3: The productivity of deadjectival-noun forming suffixes.

3.2.2. Adjective-forming Suffixes

Table 4 exhibits the overall types, overall hapaxes, and productivity values concerning the denominal adjectival suffixes *-al*, *-less*, *-ic* and *-ical* for the bases of *X-(a)(t)ion*, *X-ment*, *X-oid*, *X-(o)logy*, and *X-ist*.

	6.-al			7.-less			8.-ic			9.-ical		
base	V	n_1	P	V	n_1	P	V	n_1	P	V	n_1	P
X-(a)(t)ion	220	86	0.391	11	5	0.455	0	0	0	0	0	0
X-ment	31	8	0.258	0	0	0	0	0	0	0	0	0
X-oid	17	3	0.176	0	0	0	0	0	0	0	0	0
X-(o)logy	0	0	0	0	0	0	32	0	0	159	69	0.434
X-ist	0	0	0	1	1	1.000	137	43	0.314	12	2	0.167

e.g. 6. *gradational, managemental, sphenoidal* 8. *historistic* 9. *malacological*

Table 4: The productivity of denominal-adjective forming suffixes.

3.2.3. Verb-forming Suffixes

Table 5 presents the overall types, overall hapaxes, and productivity values concerning the deadjectival verbal suffixes *-ize*, *-ify*, and *-ate* for the bases of *X-al*, *X-(i)an*, *X-ic*, and *X-ive*.

	10.-ize			11.-ify			12.-ate		
base	V	n_1	P	V	n_1	P	V	n_1	P
X-al	106	42	0.396	0	0	0	0	0	0
X-(i)an	13	9	0.692	0	0	0	0	0	0
X-ic	14	5	0.357	0	0	0	0	0	0
X-ive	2	1	0.500	0	0	0	0	0	0

e.g. 10. *musicalize, pedestrianize, poeticize, comprehensivize*

Table 5: The productivity of deadjectival-verb forming suffixes.

3.2.4. A Suffix Sensitive to a Particular Prefix

Table 6 displays the overall types, overall hapaxes, and productivity values concerning the deverbal nominal suffixes *-ment*, *-(a)(t)ion*, and *-al* for the base beginning with *en-*.

	-ment			-(a)(t)ion			-al		
base	V	n_1	P	V	n_1	P	V	n_1	P
en-X	59	12	0.203	0	0	0	0	0	0

e.g. *enfacement*

Table 6: The productivity of deverbal-noun forming suffixes.

4. Theoretical Implications

4.1. Generalizations

Three generalizations are drawn from the results of the present investigation. To begin with, a large number of words are created by the processes concerned. A total of 774 relevant hapaxes are found in BNC, which pushes up the related productivity values. The high rate of productivity elucidates a creative aspect of each productive suffix within a specific limited domain.

Secondly, productivity is a gradient and relative concept, and accordingly it is not defined in terms of a clear-cut binary opposition of "productive" and "unproductive." Three cases of suffix selection are recognizable: (i) a suffix is automatically selected for a base form, (ii) a suffix is preferentially selected for a base form, and (iii) there is little preference in a choice between competing suffixes. The first case is well represented in Table 2. The nominalizer *-(a)(t)ion* is inevitably selected for the base form of *X-ify*; the *P* of *-(a)(t)ion* for nominalizing *X-ify* verbs is 0.369, while the corresponding *P* of *-ment/-al* is 0. Similarly, as demonstrated in Table 4, *-ical* is chosen in deriving adjectives from *X-(o)logy* nouns; the *P* of *-ical* for adjectivizing *X-(o)logy* nouns is 0.434, while the comparable *P* of *-al/-less/-ic* is 0. It should be noted here that there is a considerable difference in the degree of productivity between hapax-oriented productivity and type-oriented productivity; although the former defines *-ic* affixation to *X-(o)logy* as unproductive, while the latter may predict this process as fairly productive, since 32 word types in *X-(o)logic* are discerned in BNC.

The second case is well illustrated in Table 3. The nominalizer *-ity* has a strong affinity with the base forms of *X-able(-ible)/X-al/X-ic/X-ile/X-ar*; *-ity* is generally used in deriving abstract nouns from adjectives ending in these suffixes.[4] We also find in Table 3 that the use of *-ness* takes precedence over the use of *-ity* to nominalize *X-ous* adjectives, as evidenced by the difference in narrow productivity between both suffixes. The *-ity* or *-ness* affixation to *X-ive* adjectives, displayed in Table 3, is an example of the third case. There is little preference in a choice between these nominalizers for the *X-ive* base; the *P* of *-ity* for nominalizing *X-ive* adjectives is 0.241 and the corresponding *P* of *-ness* is 0.295.

The final generalization that can be made is that suffixation to simple words is not productive. As indicated in Tables 2 and 3, the productivity values of *-(a)(t)ion*, *-ity*, and *-ness* for nominalizing simple words are, respectively, 0.018, 0.091, and 0.090; there is no suffix which productively combines with monomorphemic bases.

4.2. Generative Morphology and the Simplified Lexicon

The generalizations sketched above substantiate the theory of "generative" morphology and its system of lexical insertion. The point of the first one was that a considerable number of hapaxes in BNC confirm a productive facet of each relevant suffix within a narrow domain. This demonstrates the thesis of generative morphology: regular complex words are constantly generated by word formation. It should be noted that experimental evidence implies that while highly frequent words are stored and easily accessible, infrequent ones are generally created by some rule (Hay, 2003: 77-81). Therefore, hapax legomena in a large corpus provide a reliable and objective indicator of a high level of productivity—the potentiality of creating new words.

Importantly, complex words produced by a productive process are not stored at all but composed by rule as needed; words such as *Mongolization* and *managemental* are aptly coined in a particular context. Jackendoff (2002: 155-159) specifically states that inventive kind of word formation like this gets involved in working memory, where items are essentially composed by free combinatory rules. It is this kind of word coinage that crucially contributes to constructing the simplified and elegant lexicon; the succinct lexicon is obtained by reducing the number of listed items

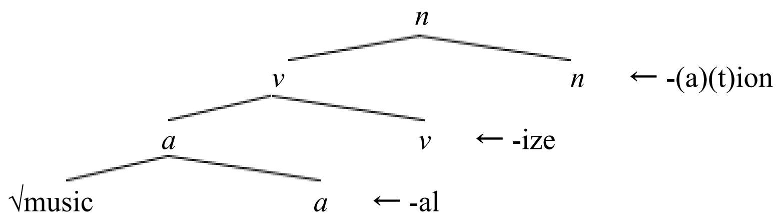
[4] A case where the number of word types is extremely low may be problematic for our productivity measure; the *P* of *-ness* affixation to *X-ic* is very high, although there are only two types of *[[X-ic]-ness]* forms attested. In relation to this, Baayen and Lieber (1991: 818-819) suggest "the global productivity *P**": the *P** of an affixation rule is defined in terms of its coordinates in the *P*-*V* interaction region, with productivity (*P*) on the horizontal axis and types (*V*) on the vertical axis; a productive affix occupies a central position in the region. According to this method, we may define a domain where productivity measurement is possible, sending the case in question outside the domain.

and generating unregistered items through word formation processes.[5]

4.3. Competition and Blocking

In a current theory of antilexicalism, Distributed Morphology (DM) (Embick, 2010), major word formation is situated within syntax: to construct a highly constrained grammar by severely restricting the morphological component and its relation with syntax, DM attributes the core properties of word construction to its syntactic structure while consigning the role of its formal make-up to the morphological module. In the DM framework, the syntactic structures of multiply affixed words are uniformly constructed by the merging of a root and category-defining heads, as shown diagrammatically in (1) (cf. Embick, 2010: 94-96).[6]

(1) *musicalization*



The second and third generalizations discussed in Section 4.1 imply that a word-form of each complex word is productively realized by Competition and Blocking. In the postsyntactic morphological module, specification of an adjectival, verbal, or nominal form is made by the addition of an appropriate affix to the base, and competition is carried out to decide which affix is chosen for a given base form in deriving a new item. In the case of nominalization, *-(a)(t)ion* wins out over its competitors for the base forms ending in *-ify*, *-ize*, and *-ate* (cf. Table 2). According to Embick's framework, the lexical entries of *-(a)(t)ion* are formalized as "*n* ↔ -(a)(t)ion /LIST 4⁀_ LIST 4={Roots, … v^{-ify}, v^{-ize}, v^{-ate} …}, where v^{-ify}, for instance, stands for *-ify* final verbs." The rival nominalizers *-ment* and *-al* are then prevented from joining to these base forms, following the narrowly defined blocking principle: competition takes place only between single morphemes (Embick and Marantz, 2008: 7); *-(a)(t)ion* competes with *-ment/-al* for an insertable nominalizer and the losers (*-ment/-al*) are blocked out. The same obtains for the verbalization of *musical* in (1) (cf. Table 5).

Another case of the second generalization is that a suffix takes precedence over its rival suffix to categorize a certain base form; *-ity* has priority to *-ness* for the *X-al* base, whereas *-ness* is prior to *-ity* for the *X-ous* base (cf. Table 3). In this case, the prior suffixal affinity is regulated in the related lexical entries, whereas the bases of the rival suffix are marked item-by-item in its entries. Thus, the lexical entries of *-ity* are "*n* ↔ -ity /LIST 7⁀_ LIST 7={Roots, [X-ous]$_1$, [X-ous]$_2$, [X-ous]$_3$, … a^{-al}, …}," while those of *-ness* are "*n* ↔ -ness /LIST 8⁀ _ LIST 8={Roots, [X-al]$_1$, [X-al]$_2$, [X-al]$_3$, … a^{-ous}, …}." The productive use of *-ity* for the *X-al* base generally blocks the addition of *-ness* to the base, although a set of *X-al* words which *-ness* takes are specified one by one in its entries.

Finally, the unpredictable bases of a suffix are specified item-by-item in its entries. Suffixation to simple words is unproductive (generalization 3) and hence which suffix preferentially combines with a given monomorphemic base is totally unpredictable. The simple base forms (roots) are then specified as in "*n* ↔ -ity /LIST 7⁀_ LIST 7={√civil, √null, √odd, √sane, …}." Another good example of this case is *-th* suffixation. Our BNC research shows that *-th* is almost always added to monomorphemic words and it takes a productivity value of 0 for this base form; ten *-th* derivatives are detected in BNC: *breadth* (575 tokens), *death* (19889), *depth* (2990), *length* (7049), *strength* (6946), *truth* (7930), *untruth* (35), *warmth* (1957), *width* (1141), *youth* (5308). Its entries can therefore be

[5] See Stemberger and MacWhinney, 1988; Frauenfelder and Schreuder, 1992 for the related psycholinguistic experiments.
[6] Root (√) is defined as bound morpheme that becomes the core of a word.

something like this: “n ↔ -th /LIST 9⌒ _ LIST 9={√broad, √dead, √deep, √long, √strong, …}.”[7]

5. Conclusion

We have proposed a new productivity measure—narrow productivity—for affixation, which crucially depends on hapax derivatives and their base forms, and then conducted an in-depth analysis of twelve kinds of derivatives identified in BNC to calculate the productivity values of relevant affixes. The results have disclosed some intriguing properties of affixation: a creative and generative aspect, competition and blocking of rival affixes, and their consequent implications for the systematic materialization of a word form. The proposed productivity measure and its consequences are expected to obtain further support by extensive research of a variety of affixes.

Acknowledgement

I would like to express my gratitude to three anonymous reviewers for their valuable comments and suggestions on an earlier draft of this paper. This work is partly supported by a Grant-in-Aid for Scientific Research (C) (No. 26370462) from the Japan Society for the Promotion of Science.

References

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Baayen, H. and Lieber, R. (1991). Productivity and English Derivation: A Corpus-based Study. *Linguistics*, 29:801-843.

Baayen, H. and Renouf, A. (1996). Chronicling *the Times*: Productive Lexical Innovations in an English Newspaper. *Language*, 72:69-96.

Embick, D. (2010). *Localism versus Globalism in Morphology and Phonology*. Cambridge, MA: MIT Press.

Embick, D. and Marantz, A. (2008). Architecture and Blocking. *Linguistic Inquiry*, 39:1-53.

Frauenfelder, U. H. and Schreuder, R. (1992). Constraining Psycholinguistic Models of Morphological Processing and Representation: the Role of Productivity. *Yearbook of Morphology*, 1991:165-183.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Hay, J. (2003). *Causes and Consequences of Word Structure*. New York: Routledge.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jespersen, O. (1949). *A Modern English Grammar on Historical Principles* VI. London: George Allen and Unwin.

Lehnert, M. (1971). *Reverse Dictionary of Present-Day English*. Leipzip: VEB Verlag Enzyklopädie.

Lieber, R. (2010). *Introducing Morphology*. Cambridge: Cambridge University Press.

Plag, I. (1999). *Morphological Productivity: Structural Constraints in English Derivation*. Berlin: Mouton de Gruyter.

Stemberger, J. P. and MacWhinney, B. (1988). Are Inflected Forms Stored in the Lexicon? In Hammond, M. and Noonan, M., Eds., *Theoretical Morphology*, pages 101-116. San Diego: Academic Press.

[7] Two types of blocking should be recognized: (i) the most productive affix for a given class of base blocks the attachment of rival affixes to the base form (class-blocking); *-ment* and *-al* are prevented from joining to $[X\text{-}ize]_V$, and (ii) when an affix is added to a given item to derive a word, the corresponding derivatives are pre-empted (item-blocking); *-ity* and *-ness* are blocked from combining with √*long*.