A Survey of the Measurements of Morphological Productivity

Lifang Du & Xuezhong Zhang

Foreign Language School, Dalian University of Technology

Dalian 116024, China

Tel: 86-411-8471-8810   E-mail: dulifang1984@gmail.com

Abstract

Morphological productivity is one of the key issues in the study of derivational morphology. This paper makes a survey of the quantitative measurements of morphological productivity so far proposed by different scholars, and tentatively attempts to point out the pros and cons and also feasibility of each measurement, with a view to provide some assistance for the future researcher who is going to carry out the study in this field.

Keywords: Morphological productivity, Type, Token, Hapax

1. Introduction

Morphological Productivity is central to the study of word-formation. It means different things to different people. The various views can be outlined as follows, as Bauer(2001, p.12) puts it in his book *Morphological Productivity*: a) Affixes are productive: The property of an affix to be used to coin new complex words is referred as productivity.(Ingo Plag,2003,p.44;Lulofs,1835,p.157,cited in Schultink, 1992a,p.189;Fleischer,1975,p.71). b) Morphological processes are productive: A property of the morphological process to give rise to new formations on a systematic basis.(Ingo Plag,2004;Uhlenbeck 1978,p.4;Anderson,1982,p.585).c)Rules are productive (Aroboff,1976,p.36;Zwanenburg,1980, p.248;Bakken,1998,p.28).d). Words are productive: (Saussure, 1969, p.228). Though there is a disagreement in the literature as to what it is that is productive, the quantitative study of the productivity is mostly centered on the affix' productivity. Various quantitative measurements have been proposed by different scholars, of which some are testing the past productivity, while some assessing the potential productivity. It should be pointed out that productivity is a diachronic phenomenon, which means that for a certain affix, it might be very productive in the past to produce a great many words, like –*ment*; however, nowadays hardly any new words are coined by using this suffix, thus becoming unproductive. This paper makes a survey of the quantitative measurements of morphological productivity so far proposed by different scholars, and tentatively attempts to point out the pros and cons and also feasibility of each measurement, with a view to provide some assistance for the future researcher who is going to carry out the study in this field.

2. Measurements of Morphological Productivity

2.1 Aronoff'model

Aronoff (1976, p.36, cited from Bauer, 2001) attempts to calculate the ratio of the actual words produced by a word-formation rule (WFR) to potential words produced by that rule. His belief behind this is that 'count up the number of words one feel could occur as the output of a WFR(which one can do by counting the number of possible bases),count up the actually occurring words formed by that rule, take the ratio of the two and then compare this with another WFR'. The formula is given as bellow:

$$I = \frac{v}{s}$$

Where *I* is the index of productivity; *v* is the number of actual/attested types, and *s* is the number of the types which the WFR could give rise to. Theoretical as well as practical problems have been pointed out by Bayeen and Lieber (1991, p.803; cited from Bauer, 2001, p.145): Firstly, the identification of the number of the existing types of a WFR is problematic. Even though V could be identified in some fixed corpus, it's not always clear whether the corpus is exhaustive of all actual types; secondly, the number of the potential bases of a WFR is hard to define given the various restrictions on the bases. In terms of many problems encountered when using this model to calculate the productivity of a WFR, It's practically unfeasible and therefore rejected in this paper.

2.2 Frequency models

Frequency models are based on the assumption that frequency is related to productivity, either directly or indirectly. The term 'frequency' means the number of times that a word occurs in a corpus. Three different models concerning

frequency will be introduced briefly, including type frequency, token frequency and relative frequency.

2.2.1 Type Frequency

People seem to hold the view that an affix is much more productive if a large amount of words were produced by using it. Therefore, one of the most widely used measure in literature is a straight count of the number of the attested types (the number of different words)with that affix at a given period of time, so is the name given as 'type frequency'. This measure is also at the same time mostly rejected by scholars, for the reason that an affix may give rise to many words in the past, but nowadays people may seldom use it to produce new words. An example of such an affix suggested by scholars is the suffix –*ment*, which in early centuries gave rise to many new words, and many of them are still in use at present, but today's speakers hardly employ it to create new words, so it would be considered rather unproductive. (Ingo Plag, 2003, p.52) However, the author holds that this measure would be better labeled as testing the past productivity of an affix at a given point of time.

2.2.2 Token Frequency

Since type frequency has its disadvantage in measuring productivity, another way to view the degree of productivity is to take token frequency into account. Bayeen (1993, cited from Ingo Plag, 2003) discussed the relationship between frequency and productivity. His main ideas are outlined as follows: A productive morphological process is characterized by a preponderance of words with rather low-frequency and a small number of high-frequency words, whereas unproductive processes with a large number of high-frequency words and small number of low-frequency words. This seems puzzled logically, however, the reasoning behind this is that: high-frequency complex words (e.g. *disadvantage* with a frequency of 1127 in BNC) are likely to be stored as whole words in the mental lexicon, and low-frequency complex words are to be stored with its decomposed parts. The reason that a newly-coined complex word (e.g. *dis-represent* with a frequency of 1 in BNC)can be understood by people who never encountered before is that people are more inclined to decompose the word into its parts, compute the meaning of its constituent morpheme, and then infer the meaning of the complex word. If this decomposition process is repeated over and over again, the representation of the affix is strengthened and made it much more readily to form new derivatives. On the contrary, for a process with large number of high-frequency words, the retrieval of the words from the mental lexicon follows the whole-word route,[1]which will not strengthen the representation of the affix, thus make it less likely to combine to other bases to form new derivatives (Ingo Plag,2003,pp.48-55).

2.2.3 Relative Frequency

Relative frequency takes into consideration the frequency of both the derived and the lexical bases, with the assumption that a process is more productive if comparatively the frequency of the derived is less than the frequency of the lexical bases, otherwise it's less productive. The explication for this is again related to the whole word access. Any reader interested in this issue can refer to Hay and Baayen (2002, p.204; 2003, pp.102-4) who gave detailed elucidation. This measure is to divide the frequency of the derived by the frequency of the lexical bases. The higher the figure, the less productive the process or the affix is. However, when in practical application several methodological problems arise. Firstly, how to calculate the frequency of lexical bases of a morphological process, since it's not so easy a question to sum the number of the frequency of each lexical base, let alone how many bases there will be. Secondly, can this measure authentically reflect the truth even if the above methodological problem can be cleared away? What if the following case is presented: for some words with the given affix, the derivatives are more frequent than the bases, while for others, the derivatives are less frequent than the bases? Therefore, this measure need further to be improved and developed.

2.3 Probabilistic model

Probabilistic models were proposed mainly by scholar Baayen. (1989; Baayen and Lieber 1991, p.819; 1992; cited from Bauer, 2001, p.154).The set of models statistically measure the probability of encountering a new word by a given morphological process. In probabilistic models, the calculation of the productivity is indispensably involved a crucial facor-hapax legomema(or hapax for short).Hapax are words that occur only once in a corpus. According to Baayen, if one wants to study the productivity, then it's important to study hapax.One may ponder to ask why, and what is the relationship between productivity and hapax? Plag (2003, p.54) suggested the reason that "…the number of hapaxes of a given morphological category should correlate with the number of the neologisms of that category, so the number of hapaxes can be seen as an indicator of productivity. Though Bauer(2001,p.150)raised the doubt that why hapaxes in a corpus should correspond in any meaningful way to coinages in real use, for Inevitable is that in a corpus there exist some hapaxes out of tag errors and misspelling. Anoroff discussed the importance of hapaxes in the book *What is Morphology?* He pointed out that the hapaxes in a corpus are more likely to have been formed by a productive process. The writer goes along with the view, thinking that hapaxes are mostly produced unconsciously by community members following some morphological rules, accordingly, large amount of the hapaxes of a given morphological category can indirectly indicate the productivity, even though the following possibility can not be eliminated that the corpus is too

small to include some hapaxes which are actually common words and familiar to the community members. Therefore, in order to make sure the accuracy of the results, the corpus would better be large enough. The probabilistic models have three phases, each of which will be tackled briefly in the following.

2.3.1 The first phase: P in the narrow sense (Baayen, 1989).The formula is as follows:

$$P = \frac{n_1}{N}$$

Where P is the index of productivity; n_1 is the number of words formed by the appropriate process occurring only once (the hapax) and N is the total token frequency of words created by that morphological process in the corpus.

2.3.2 The second phase:

Since the first phase doesn't take type frequency into account, Baayen (1989; Baayen and Liber 1991, p.817ff; Baayen 1992, pp.122-125; cited from Bauer, 2001) in this phrase reintroduces this in a measure of 'global productivity.' He adopts a two-dimensional chart to show the productivity of a given affix, with the horizontal axis indicating the P in the narrow sense, and the vertical axis indicating the type frequency, see the following figure:

Insert Figure 1 Here

From the chart, one can have a visual impression of productivities of different morphological processes. Those dots located in left-bottom corner show the lower productivity, while that in top-right hand show higher productivity. However, this measure still could not escape the fate of objections by some scholars, even Baayen (1992, p.24) admits, it's not possible to weight the relative contributions of the vertical and horizontal dimensions in such a chart. In view of this, Bayyen (1993, p.192) proposes yet another measure which he terms 'the hapax-conditioned degree of productivity.

2.3.3 The third phrase: the hapax-conditioned degree of productivity. Baayen formulizes it as:

$$P *= \frac{n_{1,E,t}}{h_t}$$

Where E indicates the appropriate morphological category, t indicates the number of tokens in the corpus and h is the number of hapaxes. This measure computes the ratio of the number of hapaxes with a given morphological category with the total number of the hapaxes in the corpus. Since the denominator (total number of the hapaxes in the corpus) is a constant value, the P* value is dependant on the hapaxes of the given morphological category. This measure tests 'expanding productivity' (Baayen, 1992), while 'P in the narrow sense' is labeled as testing the potential productivity.Baayen gives an interesting metaphor to show the difference between the two productivities. A rule that is productive in the first sense is like a company that is expanding on a market (no matter whether the company has a large share of the market or not. A company may have a large share of the market, but if there are hardly any buyers because the market is saturated, the company is in danger of going out of business, so the measure **'P in the narrow sense'** gauges the extent to which the market for a category is saturated.

Apart from the measures outlined above, some other measurements are proposed by scholar Stekauer, which he terms as 'the onamasiological model.' This measure is distinct from the above in that it goes from meaning to form rather than from form to meaning. For more about this measure, the readers who are interested can refer to Stekauer (cited from Jesús Fernández-Domínguez, 2007).

3. Conclusions

Scholars have been trying to provide an effective way of accessing the productivity of affixes quantitatively. However, it seems that no one of the measurements is hardly without any objections, either theoretical or in practical application. Nevertheless, those varied measures could be seen as showing productivity from different aspects; they are more taken as indicating the productivity in a comprehensive and multi-angle point of view rather than contradictory to each other.

References

Aronoff, Mark & Fudeman, Kirsten. (2005). *What is Morphology?* Blackwell publishing.

Baayen, R.Harald. (1992). Quantitative aspects of morphological productivity. In: Greert. E.Booij & Jaap. Van Marle (Eds).*Yearbook of Morphology 1991*.Dordrecht: Kluwer Academic Publishers, 109-149.

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

Jesús Fernández-Domínguez. (2007). How is Morphological Productivity measured? *Atlantis* 29.1(June 2007)29-54.

Matthews. (2000). *Morphology*. (2nd. Ed) Beijing: Foreign language teaching and research press.

Plag, Ingo. (2003). *Word Formation in English*. Cambridge: University Press.
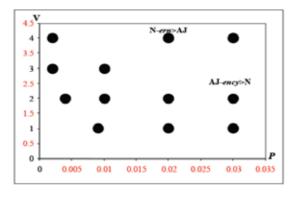
Figure 1. Global Productivity of a Number of English Word-formation

(This chart is taken from Jesús Fernández-Domínguez, 2007)