ATLANTIS 29.1 (June 2007): 29–54
ISSN 0210-6124

How is Low Morphological Productivity Measured?

Jesús Fernández-Domínguez
Universidad de Jaén
jesusferdom@gmail.com

Ana Díaz-Negrillo
Universidad de Jaén
adinegri@ujaen.es

Pavol Štekauer
Pavol Jozef Šafárik University in Košice
stekauer@condornet.sk

Morphological productivity has become a central issue in present-day English word-formation over the past decade. However, most proposals for assessing productivity have focused on the most productive processes and how to measure them, to the detriment of processes which give rise to fewer items than the ones usually studied in analyses of word-formation productivity. The present paper looks at the major models for productivity measurement and applies them to a BNC-based 5,891-item corpus to test how they account for low productivity. The results obtained highlight significant differences between various productivity counts in an area which, from this point of view, needs further methodological development.

Key words: English morphology, present-day word-formation, low productivity, productivity measurement, constraints, corpus-based research.

1. Introduction

Defined as "… the possibility for language users to coin, unintentionally, a number of formations which are in principle uncountable …" (Schultink 1961), morphological productivity has been neglected for decades. Today this topic attracts the attention of many scholars, so much so that a considerable number of publications have been made in the field over the last decade (Baayen 1992, 1993, 1994, 2003; Bauer 2001, 2005a, 2005b; Kastovsky 2005; Plag 1999, 2003, 2006; Štekauer 1998, 2001, 2005a, 2005b, 2006).[1]

[1] Acknowledgements are due to Dr. Salvador Valera Hernández, University of Jaén, for his elucidating commentaries on the content and formal aspects of this paper. We are also grateful to the three anonymous reviewers who offered a number of useful suggestions to give this paper its final shape.

The above-mentioned growth in studies of this type has meant an increase in proposals for productivity measurement, most of them with a primary focus on the study of high productivity. Using the models available usually applied to the study of high productivity, this paper examines low productivity, and specifically, aims to explore:

i.) the implementation procedures of existent productivity measurements for the study of low productivity,

ii.) corpus-based low productivity areas according to each model, and

iii.) the potential of each model on the grounds of low productivity measurement.

To this end, six models of productivity measurement have been selected, often preferred for their simplicity of application (see 4.1.1 and 4.1.2), or ability to predict future formations (see 4.2) (*cf.* Bauer 2005a). They have been applied to a 5,891-item sample of the *British National Corpus* (hereafter BNC), after which the ten least productive affixes from each model are extracted and studied as representative of low productivity with respect to the points mentioned above.

Following this introduction, section 2 reviews theoretical grounds for low morphological productivity on the basis of productivity constraints. The methodology is described in 3, and the major models for productivity measurement are tested and assessed in 4. Section 5 discusses the results and other remaining issues concerning productivity measurement, while 6 offers some conclusions.

2. Constraints on Productivity

While there are factors that seem to favour productivity, such as *semantic coherence* or *naturalness* (*cf.* Aronoff and Anshen 2001: 246; Bauer 2001: 20; Kastovsky 1986: 586), the coinage of a word also has to overcome difficulties at the levels of both language use and language structure before it becomes materialized. In the remainder of this section, some of the constraints responsible for low word-class productivity are arranged for review under structural constraints, pragmatic constraints and blocking.

2.1. *Structural Constraints*

Structural constraints relate to the form of items at various levels of description, namely phonology, morphology, syntax and semantics. Phonological constraints are usually associated with cacophony and the possible ill-formedness of the potential word (**elderlily, *wordlily*), but other types of phonological constraints are mentioned in the literature (Bauer 2001: 128-29; *cf.* also Aronoff and Fudeman 2005: 216; Giegerich 1999: 3-5; Katamba 1993: 74-75; Rainer 2001: 881; 2005a: 344-45; Yip 1998) on the grounds of:

i.) the segmental constitution of the word, as in (1), where the suffix *-en* is appended only to words ending in a stop or a fricative:

(1) *neaten, quieten, smarten* and *tighten* (Marchand 1969: 213)

ii.) its suprasegmental structure, as in (2), where the suffix *-al* is appended only to verbs stressed on the final syllable:

(2) *arrival*, *rebuttal* (Bauer 2001: 129)

The morphological structure of the base may also condition a potential formation (Bauer 2001: 130-31; *cf.* also Aronoff 1976: 51-63; Aronoff and Fudeman 2005: 217; Katamba 1993: 76-77; Rainer 2001: 881-82). Two well-known particular conditions for a formation to be produced are:

i.) the base has to belong to a morphologically-defined class. Example (3) shows how verbs ending in *-ize* accept nominalization by *-ation*, but not by other suffixes:

(3) *colonization* vs. **colonizement*, **colonizal*, **colonizage* (Plag 2006)

ii.) either there has to be a specific affix or not, as in (4), where the denominal suffix *-ity* is attached to bases not ending in *-ory*:

(4) *polarity*, *peculiarity*, *scalarity* vs. **notorious-ity*, **adventurosity* (Plag 1999: 88-89)

In syntactic constraints, the focus is on the restriction of word-formation processes to members of certain syntactic categories (Bauer 2001: 133; Plag 2003: 63). For instance, the prefix *un-* is largely limited to adjectives and some verbs, and the suffix *-able* is attached only to transitive verbs, as in (5):

(5) *comparable* (Plag 1999: 56) vs. **becomable* (Bauer 2001: 133)

Finally, in semantic constraints the concern is the referent of a given word, such that there can be a limitation in relation to what things there should be a word for.[2] A classical example for this is adjectives ending in *-ed*, where "the base must be inalienably possessed by the head noun that the adjective modifies" (Quirk *et al.* 1985: 1329; *cf.* also Katamba 1993: 78; Rainer 2005b):

(6) *blue-eyed*, *three-legged*, *red-roofed* vs. **a black-shoed lady*, **a two-carred man* (Quirk *et al.* 1985: 1329)

Despite the ruling role they generally play, it should be noted that constraints are rarely infallible. Exceptions to constraints have been mentioned by Bauer (2001: 130),

---

[2] An important distinction is made by Bauer (2001: 133) between semantic and lexical constraints: the latter refer to the fact that, in non-productive affixation, the stock of words to which a given affix can be added is limited, often due to historical constraints (cf. Plag 1999: 35, 45), as in *bishopric* (Bauer 2001: 135, showing the suffix *-ric*) and *laughter*, *slaughter* (Bauer 2001: 135, showing the suffix *-ter*).

who cites *sillily*, as a contradiction to the phonological constraint above mentioned (*cf.*, however, Plag 1999: 47 for the opposite view).

2.2. Pragmatic Constraints

Pragmatic constraints may also affect the coinage of a neologism, and impose conditions on the possibility of the coinage. Some pragmatic constraints are:

i.) there has to be a need, otherwise the item would be redundant for language use (Kastovsky 1986: 595; Lieber 2004: 96; Plag 1999: 39), as illustrated in (7):

(7) "…and whether your own conversation doesn't sound a little *potty*. It's the *pottiness*, you know, that's so awful" (Kastovsky 1986: 595, italics as in the original)

ii.) the object denoted must be nameable, i.e. something we can think of when we hear the word. For instance, it would be improbable to have a verb-forming category that creates items with the following meaning:

(8) grasp NOUN in the left hand and shake vigorously while standing on the right foot in a 2.5 gallon galvanized pail of corn-meal-mush (Rose 1973: 516)

iii.) the object of the label needs to exist, so only existing things can be named. This constraint has been referred to as *Hypostasierung* ('hypostatization') (Lipka 1977: 161-62), as shown in (9):

(9) *time-machine*, *warp speed*, *beam me up* (Hohenhaus 2005: 356)[3]

Finally, *fashion* – or *aesthetics* – has also been associated with pragmatic constraints. It is said to bear influence on the use of certain prefixes or suffixes which people prefer to make use of at some time or another according to language trends, as those in (10):

(10) *mega-*, *giga-*, *supra-* (Plag 1999: 39)

Alongside structural and pragmatic constraints, a third type is sometimes cited, *sociolinguistic constraints*, which inhibit a person's creating a new word. The studies made on sociolinguistic constraints often intertwine with pragmatic constraints, so little can be discerned about this at this stage (*cf.* Bauer 2001: 135; Plag 2003: 60; Quirk *et al.* 1985: 1531; Rainer 2001: 883).

[3] Plag (1999: 40) notes that *existing* here means not only happening in the real world strictly speaking, but also *fictional existence*, i.e., "any new derivative must have some kind of referent or denotatum". An exception to the rule, however, occurs when a word is created before the denotatum exists, i.e., the speaker first invents a word and then the object referred to by that word (Plag 1999: 39-40; cf. Hohenhaus 2005: 356), frequent in child language.

2.3. Blocking

Also known as *Avoid Synonymy Principle* (Kiparsky 1983: 15), blocking explains the non-occurrence of a word due to the existence of an item with the same meaning or form. Blocking does not avoid the coining of a word itself, but rather its institutionalisation among speakers, i.e. its wide usage in the community (Bauer 2003: 80-81). This implies that a new word may appear in the language, be used for a short time, and then rapidly disappear in favour of a previously existing one.

Two categories have been distinguished in relation to blocking: *homonymy blocking* and *synonymy blocking* (Bauer 2001: 136; *cf.* also Plag 1999: 50, 2003: 64; Spencer 1991: 89). *Homonymy blocking* explains the non-occurrence of a new formation when a formally identical form already exists; here, the two items would overlap formally and cause ambiguity, as in (11):

(11) ?*liver* vs. *liver*
'someone who lives' 'inner organ' (Plag 1999: 50)

In turn, s*ynonymy blocking* is one of the devices language uses to avoid exact synonyms: a potential item is blocked, if there is already an existing item in the language which denotes the same reality. A classical example of this is (12):

(12) *thief* vs. **stealer* (Plag 1999: 50)[4]

Token-blocking and *type-blocking* have sometimes been described as subtypes of synonymy blocking (Bauer 2001: 137-38). The former refers to the blocking of potential words by actual words, and is influenced by synonymy, productivity and frequency (the more frequent the word, the more likely the blocking of a potential item). Token-blocking is the most common type of blocking and affects the profitability of word-formation processes (Aronoff 2001: 347; Aronoff and Anshen 2001: 240; Bauer 2001: 137, 2003: 80-81; Plag 1999: 51, 2006: 126, 2003: 67-68; Rainer 2005a: 336-37).[5] By contrast, type-blocking affects word-formation processes. It takes place when one process blocks another, and thus prevents the creation of new words. For example, the suffix -*ness* blocks the suffix -*ity*, for both create deadjectival nouns, and the process can be applied only once. Since it does not affect individual items, frequency does not play a relevant role here, which means that type-blocking is influenced only by synonymy and productivity.[6]

[4] Synthetic compounds are however not equally affected (cf. *sheep-stealer* vs. **stealer*, Bauer 2003: 81).

[5] *Profitability* stands in opposition to *availability* (Corbin 1987: 177) and is one of the basic concepts in productivity studies. The former is the extent to which a process is productive or not, while the latter refers to whether the process can be used or not. Profitability is a matter of degree, while availability is a yes/no question (cf. Plag 2006: 122).

[6] In the literature, although type-blocking has been studied by some authors (Aronoff 2001: 347; Bauer 2001: 138; Rainer 2005: 337-39), others have rejected the notion on the grounds that it "...rests on false assumptions about the meaning of putatively rival affixes and that it cannot account for the empirical facts" (Plag 2003: 67-68).

3. Data Preparation

The study sample used for this paper is made up of derived words in present-day English as attested in:

i.) ten BNC frequency lists (Leech, Rayson and Wilson 2004), one for each of the main word-classes, namely noun, verb, adjective, adverb, pronoun, determiner, determiner pronoun, preposition, conjunction and interjection,

ii.) *The Compact Edition of the Oxford English Dictionary* (Murray *et al.* 1971), *A Supplement to the Oxford English Dictionary* (Burchfield 1987) and *The Oxford English Dictionary* (Simpson and Weiner 1989; all cited hereafter as *OED*); this was also used to verify directionality in cases of conversion, and

iii.) a BNC-based frequency list (Kilgarriff 1996) with a frequency range from 6,187,267 to 1; this was also used to retrieve the lexical bases of derived words and hapaxes[7] necessary for application of some productivity measurements (see 4.2).[8]

The 5,891 words in the lists cited under i) above were screened for irrelevant units, after which 2,538 items remained. This process excluded records which were apparently not actual words *(£1*, ***base/basis*, *50%*), multi-word entries (*according to*, *in the light of*), items which are formally identical today but which come from formally different Old English (hereafter OE) words (*fire*, *murder*), and words which seemed morphologically simple, but which the *OED* proved to be complex (*biology*, *capacity*). Extraction of these units and others pertinent for the analysis was made manually.

Once all morphologically complex items were filtered, they were analysed and tagged accordingly. Tags were tailor-made and provided the following information: input word-class, affix involved and output word-class.[9] The illustration below shows that the affix *-ish* is appended to nouns to create adjectives:

(13) N-*ish*>AJ

Tagging was done manually and served for the retrieval of varied items (not only the lexically-dependent), and for easier examination of the tendencies of the affixes studied in terms of the word-classes involved. The labels used are shown in Table 1:

---

[7] The notion of hapax legomena is used in various senses in the literature, since certain items can be argued to be hapaxes or not. The term is here used in the general sense and refers to words with frequency 1, independently of any other relationship. For instance, if *disposal*, *refusal* and *survival* had frequency 1 and were the only elements including the suffix V-*al*>N in the study corpus, the value for hapax legomena for such suffix would be 3 (cf. Bauer 2001: 151; Plag 1999: 28-29; Plag et al. 1999: 215; Štekauer 1998: 32; Tournier 1985: 404).

[8] Kilgarriff's (1996) list was used as a reference for BNC frequencies.

[9] Such tagging is in accordance with the *Unitary Base Hypothesis* (cf. Aronoff 1976, Scalise 1984), which argues that affixes do not attach to any type of word-class, but opt for only one specific category. For example, if *-able* can attach to verbs (*acceptable*) and to nouns (*charitable*), this hypothesis requires recognition of two homophonous affixes, one V-*able*>AJ and another N-*able*>AJ.

Code	Word-Class
AJ	adjective
AV	adverb
D	determiner
N	noun
O	ordinal
V	verb

Table 1. Word-class tags

4. Productivity Models

The study sample was used for the implementation of the major models of productivity measurement, here grouped under:

i.) frequency models, which cover type frequency, token frequency and relative frequency,

ii.) probabilistic models, which cover productivity in the narrow sense, global productivity and the hapax-conditioned degree of productivity, and

iii.) the onomasiological model.

4.1 Frequency Models

Frequency models rely on the assumption that frequency of occurrence is related to productivity, either directly or indirectly. In all the three cases considered here, frequency procedures are applied in the literature reviewed only to items formed by affixation.

4.1.1 Type Frequency

This model is widespread in morphological productivity studies and is based on the concept of *type*, customarily defined as each different word that has been coined by means of the word-formation process under study (*cf.* Bauer 2001: 47-49; Plag 2003: 52-55). Accordingly, the proposed procedure is to sum up the number of types containing the relevant process, so that the higher the figure obtained, the more productive the process is. The following are the ten least productive affixes in our corpus according to type frequency:

Affix	Frequency
1. AJ-*ancy*>N	1
2. AJ-*er*>N	1
3. N-*ful*>N	1
4. N-*hood*>N	1
5. AJ-*ian*>N	1
6. N-*ish*>AJ	1
7. N-*ive*>AJ	1
8. V-*tion*>N	1
9. *un*-AV>AV	1
10. *up*-N>AV	1

Table 2. The ten least productive affixes for the type frequency model[10]

As has been mentioned, the study was undertaken only on affixated items: in this case 902 affixated types out of 1,468 complex types. One outstanding feature of the affixes in Table 2 is that they are all derivational (*cf.* Baayen and Lieber 1991: 823). Thus, the word-class of the bases involved are adjective, adverb, noun and verb. Out of these ten affixes, six form nouns, which to a certain extent can be representative of real language, where nouns are the most frequent word-class (the word-class noun is the most frequent in Leech, Rayson and Wilson's 2004 list with 3,030 units, followed by verb with 1,112 units). Suffixation surpasses prefixation in these results, with eight suffixes and two prefixes.

This model has been usually rejected in the literature on the grounds that it cannot relate past productivity to future productivity (*cf.* Bauer 2001: 48-49; Hay and Baayen 2002: 2 *et passim*; Plag 1999: 11; Štekauer *et al.* 2005: 12-13), and it cannot tell about the availability of the process under study. Additionally, Aronoff (1976: 36; *cf.* Baayen 1992: 110-11) explains that affixes cannot be compared simply by contrasting their frequency, because not all affixes are freely attachable to a base (see 2).

4.1.2 Token Frequency

In this case the basis for computations is a *token*, which here is a synonym for *occurrence*. As in type frequency, an estimate is obtained by summing up all words containing a given affix; accordingly, the higher the figure obtained, the higher the productivity. Token frequency seems to measure more variables than type frequency since it is able to display a more refined picture of the occurrences (Baayen and Lieber 1991: 804-5; Bauer 2001: 147; Hay and Baayen 2002: 204; Plag 2003: 205). Here the

[10] Please note that since 33 affixes had frequency 1 in the list of results, the number of affixes taking part in the study is greater than in other cases. The ten affixes shown in Table 2 have been randomly selected for illustration, while all affixes with frequency 1 are displayed in Table 11.

total number of complex tokens in the corpus is 59,900, and the number of affixated tokens is 32,835. Among them, the following are the ten least productive:

Affix	Frequency
1.*off*-V>V	10
2. *un*-AV>AV	10
3. *infra*-N>N	10
4. AJ-*ism*>N	10
5. N-*y*>N	10
6. *under*-N>PR	10
7. AJ-*cy*>N	11
8. *in*-N>N	11
9. *out*-N>AJ	11
10. AJ-*er*>N	12

Table 3. The ten least productive affixes for token frequency

Here the number of prefixes is larger than the number of suffixes, which is remarkable considering that our study included in its computations 91 different suffixes and 33 prefixes.[11] Some of the prefixes shown do not seem, a priori, highly productive, such as *under*-N>PR (which creates closed word-class lexemes, as in *underneath*) or *out*-N>AJ (as in *outdoor*). However, the high figures they display may be justified by the high frequency of the derived units in which the affix appears. So, although token computations may disclose more refined results, they may also be ambiguous, which is a major drawback of this model (*cf.* Bauer 2001: 147; Lyons 1977: 20; Plag 2003: 50). Other major criticisms coincide with those levelled at the type frequency model, and relate to its great reliance on frequencies and its incapacity to determine availability and to explain future or past productivity (*cf.* Bauer 2001: 48-49; Hay and Baayen 2002: 2 *et passim*; Plag 1999: 11; Štekauer *et al.* 2005: 12-13).

4.1.3 Relative Frequency

Unlike type and token frequency, relative frequency takes into consideration the frequencies of both the derived word and its lexical base, and maintains that a word-formation process is more productive when the derived items are less frequent than their lexical bases. An explication for this is found in the concept *whole word access* (Hay and Baayen 2002: 204, 2003: 102-4), which explains that, in cases where the derived word is more frequent than its base, the speaker tends to see the formation as

[11] Only the elements regarded as affixes in the OED were considered as such for this study. Once the affixes were retrieved, the word-class tags appended (see Table 1) were used for identification and analysis (see 3).

an indecomposable entity, which may ultimately contribute to the unproductivity of the affix. In this case, figures are obtained by dividing the frequency of the derived word by the frequency of its lexical base, that is, an inversely proportional operation such that the lower the results, the higher the productivity. Table 4 shows the ten least productive affixes in our corpus according to relative frequency:

Affixes	Figures
1. AJ-*ian*>N	133.30
2. AJ-*ic*>AJ	30.60
3. V-*ion*>N	13.50
4. N-*ar*>AJ	12.12
5. V-*our*>N	11.68
6. N-*ish*>AJ	6.90
7. N-*y*>N	6.71
8. N-*ish*>AJ	5.79
9. V-*ee*>N	5.38
10. *under*-N>PR	4.50

Table 4. The ten least productive affixes for relative frequency

Here, nine out of the ten affixes form either nouns or adjectives and, except for *under*-N>PR, all of them are related to open word-classes, both in their base and in their derived term. In this sense, results according to relative frequency seem to closely reflect the linguistic reality in that the number of words in our corpus with those affixes is relatively low; for instance, 5 for V-*ee*>N, 3 for AJ-*ic*>AJ, 1 for AJ-*ian*>N, or 1 for *under*-N>PR.

In compounds, problems of measuring are encountered in the computation of the frequency of the base, since there are various possibilities: i) summing up the frequencies of the separate constituents, ii) summing up the frequencies of the separate constituents and dividing them by the number of constituents, and so calculating the average, or iii) using only the frequency of the head. As shown (Fernández-Domínguez 2006), each possibility yields different results, and thus may raise doubts about the consistency of the present model.

4.2 *The Probabilistic Models*

Probabilistic models rely largely on Baayen's research (1992a, 1992b, 1994, Baayen and Lieber 1991), which develops from Aronoff's principle that token frequency is related to semantic complexity. Thus, token frequency is central to Baayen's models. Specifically, his computations are based on the notion of *hapax legomena*. The relationship between hapaxes and productivity is that, in a productive word-formation process, there are more words with low frequency (such as hapaxes) and fewer with high frequency. In unproductive processes, the tendency is the opposite: there are more items with high frequency and fewer items with low frequency, hapaxes among them. The justification for this is that, if a process is productive and creates a large number of new lexemes, their frequency is distributed among all of them, thus resulting in a lower figure. If the process is unproductive, the same items will be used more regularly, so their frequency value will be higher. This model, then, assumes a correlation between the number of hapaxes and the number of neologisms, which over time is an indicator of productivity. As in the frequency models, Baayen's models deal only with affixation.

4.2.1 P *– Productivity in the Narrow Sense*

The first model is called *productivity in the narrow sense*. Here, Baayen and Lieber (1991: 809; see also Baayen 1992: 115-16, 1994: 450-51; Hay and Baayen 2003: 101) propose an indirect relationship between the number of hapaxes and the token frequency in a given affix, and express it with the formula:

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

where P is the measurement of productivity, n_1 is the number of hapaxes that contain the affix under study and N is the token frequency of all items with that affix. Hence, the higher the value of P, the higher the probability of finding a new coinage with a given affix (*cf.* Bauer 2001: 147-48, 2005a: 325-26; Plag 2003: 56-59). The following are the ten least productive affixes in our corpus according to P:

Affix	Figures
1. *down*-N>AV	0.001
2. N-*ern*>AJ	0.002
3. N-*ly*>AJ	0.002
4. *en*-AJ>V	0.002
5. N-*ful*>AJ	0.003
6. *cor*-N>N	0.003
7. V-*ment*>N	0.004
8. AJ-*wise*>AV	0.004
9. *dis*-V>V	0.004
10. N-*ship*>N	0.005

Table 5. The ten least productive affixes for P

In some cases, the existing words with the affix under study are certainly very few, as in N-*ern*>AJ, which is mainly restricted to *eastern*, *western*, *southern* and *northern*, or *down*-N>AV, in our corpus represented by *downstairs*. So it could be said that P is quite accurate in reflecting language reality. Similarly, no inflectional affix occurs in the list, which would agree with the general tendencies stated in the literature (*cf.* Baayen and Lieber 1991: 823; Roeper 2005: 126). The methodological procedures have raised some criticism, for instance Plag (1999: 28), who explains the difficulties in the computation of hapaxes. Some remarks on the foundations of the model are Baayen and Lieber's (1991: 816-17), which admit that the model cannot account for potential formations, i.e. future productivity; this has been alleged to provide an incomplete picture of productivity (Štekauer *et al.* 2005: 5). To Baayen and Lieber's shortcoming, Bauer (2001: 151-53) adds the impossibility of the model to determine the availability of the process. Finally, Baayen (1992b: 117) also acknowledges the variability of *P* with respect to the size of the corpus as a further disadvantage.

4.2.2 P – Global Productivity*

Baayen's second proposal is *global productivity*, symbolized P*, where V, type frequency, is computed together with the figure of productivity in the narrow sense, *P*. Thus, P* is calculated in a two-axis chart, in which the value of V is shown on the *y* axis and the value of *P* on the *x* axis. Each word-formation process is represented on a chart by a dot, and the rationale is that the closer a process is to the bottom left-hand corner, 0 in the *x* and *y* axes, the less productive it is (see Fig. 1). The display of the results in this type of chart is advantageous since it allows observation of more nuances of productivity than a single figure does (as in the frequency models, in *P* or in P^*), since others measures of productivity are also described, type frequency (4.1.1), here V, and productivity in the narrow sense (4.2.1), here *P*.

Two suffixes, N-*ern*>AJ and AJ-*ency*>N, have been chosen from Fig. 1 to illustrate this point. Here, N-*ern*>AJ has a V value 4 and a *P* value 0.02, AJ-*ency*>N has a V value 2 and a *P* value 0.03, while both suffixes seem to be at the same distance from the 0 in *x* and *y* axes. The nuances here are the following: while both suffixes seem to be equally productive for the present model, their V and *P* values show that according to their type frequency AJ-*ency*>N is less productive than N-*ern*>AJ and, according to productivity in the narrow sense, N-*ern*>AJ is less productive than AJ-*ency*>N.

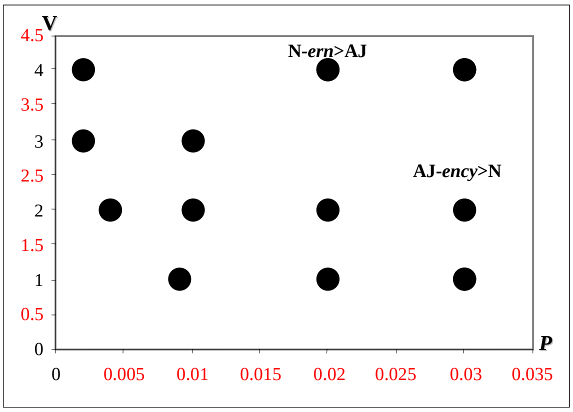

Figure 1. Global productivity (P*) of suffixes[12]

AJ-*bility*>N	N-*ful*>AJ	N-*ly*>AJ
AJ-*en*>V	N-*ful*>N	V-*our*>N
AJ-*ency*>N	N-*hood*>N	AV-*case*>AV
N-*ern*>AJ	AJ-*ically*>AJ	AV-*ward*>AV
N-*ese*>AJ	N-*ise*>V	AJ-*wise*>AV
V-*fication*>N	N-*less*>AJ	N-*y*>AJ

Table 6. Suffixes displayed by P*

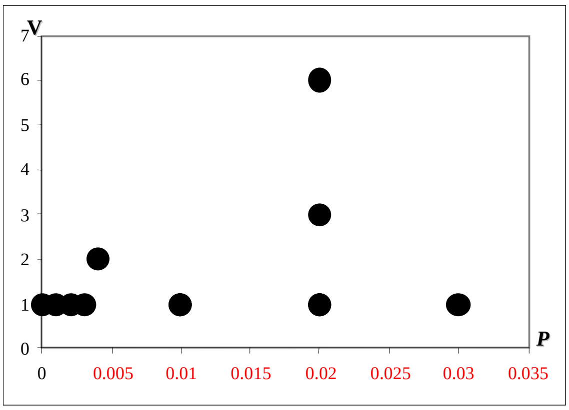

Figure 2. Global productivity (P*) of prefixes

[12] Please note that the chart illustrates productivity of the fifteen least productive affixes because the model's requirements made it objectively impossible to select the ten least productive ones. The number of affixes differs from the ten selected from other models for comparison. However, here this is not relevant since its comparison is from the start hindered by its own format.

back-N>N	*en*-AJ>V	*off*-V>V
cor-N>N	*en*-V>V	*out*-V>V
dis-V>V	*fore*-V>V	*un*-AJ>AJ
down-N>AV	*infra*-N>N	*under*-N>PR
en-N>V	*inter*-AJ>N	*under*-V>V

Table 7. Prefixes displayed by P*

While the visual display of the chart can be advantageous, as explained above, it may also stand as an important limitation when it comes to comparing it with other models where results are in figures (Baayen and Lieber 1991: 818; *cf.* Plag 1999: 32). Similarly, P* cannot be used to compare suffixes among themselves directly, because V and *P* do not allow unified calculation (Baayen 1992: 123). Besides, Bauer (2001: 154) explains that the values of V and *P* in the above charts are not significant, since their combination is arbitrary and does not provide significant results about the productivity of an affix.

4.2.3 P – The Hapax-Conditioned Degree of Productivity*

The *hapax-conditioned degree of productivity* is intended to compare the number of hapaxes from a given word-class in the corpus with the total number of hapaxes in the same corpus. The formula set for this is:

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

where n_1 is the number of hapaxes that contain the affix under study, *E* is a given morphological category,[13] and *t* and *h* are the number of tokens and the total number of hapaxes in a corpus respectively (Baayen 1994: 451; *cf.* Hay and Baayen 2003: 101). The following are the ten least productive affixes in our corpus according to *P**:

[13] The term *morphological category* is used here to mean the particular affix studied. For instance, one morphological category could be the prefix *-down* added to nouns and creating adverbs (*down*-N>AV), and another, the suffix *-ment* added to verbs to create nouns (V-*ment*>N).

Affix	Frequency
1. *under*-N>PR	0
2. *cor*-N>N	0.00001
3. AJ-*ancy*>N	0.0001
4. N-*ern*>AJ	0.0001
5. N-*ful*>N	0.0001
6. AV-*case*>AV	0.0001
7. AJ-*acy*>N	0.0002
8. AJ-*en*>V	0.0002
9. V-*fication*>N	0.0002
10. AJ-*hood*>N	0.0002

Table 8. The ten least productive affixes for P^*

The results obtained by P^* are very closely related to those by P (see 4.2.1), since both formulas list affixes with few items formed nowadays. The results in P^* include five verb-forming affixes. This is outstanding in comparison with the results in other models, where the output word-class is usually either nouns or adjectives. Besides, the number of suffixes shown by P^* is also relevant, eight out of the ten, all of which occur in low-frequency items in our corpus; for example, 18 for *cor*-N>N, 18 for AJ-*ancy*>N, 17 for AJ-*acy*>N or 12 for AJ-*hood*>N.

Bauer (2001: 155; *cf.* 2005a: 326) has some reservations about this model specifically about "… whether P^* is measuring the right thing" (2001: 155), i.e. if hapaxes can be taken as a sign of possible coinages. Moreover, for him the figures of hapaxes are not clearly relevant for the prediction of coinages (see also Plag 1999: 33).

4.3. The Onomasiological Model

The main representative figure of the onomasiological model is Štekauer, whose work on word-formation differs from the models described above in that it goes from meaning to form (onomasiological approach), rather than from form to meaning (semasiological approach). For him, word-formation processes come into play when the speaker has a specific need for a Naming Unit (hereafter NU, a synonym for *word*, *lexeme* or *lexical unit*) which is not contained in the Lexical Component, i.e. the lexicon of a language. One basic feature is that, in contrast to analysts who support the higher productivity of syntax and inflection (Chomsky 1970, Baayen and Lieber 1991, Roeper 2005), word-formation patterns are here claimed to be regular, predictable and absolutely productive (Štekauer 1998: 73-25; 2001: 6-7; Štekauer *et al.* 2005: 3-4). The former position traditionally leads to perceiving productivity as a rule-governed and limited phenomenon vs. creativity as a hotchpotch of idiosyncrasies sheltering all remaining output of language. Such a view is automatically declined in the onomasiological approach where, instead, both notions are related by a conception of word-formation as "...creativity within productivity constraints" (Štekauer *et al.* 2005: 15).

Štekauer *et al.* (2005: 15-16) illustrate the above point for better understanding: if twenty people are asked to provide a new NU for "a person who meets space aliens on behalf of the human race", it will rarely be the case that all twenty participants utter the same word, and they will probably supply a range of NUs with different morphological realizations. Table 9 presents a number of possible NUs for the above referent together with their corresponding onomasiological word-formation type:

Word-formation type	Possible morphological realizations
Theme – Action – Agent	*human race representative*, *homosapience representative*
Location/Theme – Action – Agent	*earth-representative*, *world ambassador*
Location – Action – Agent	*intergalactic diplomat*, *interstellar diplomat*
Object/Location – Action – Agent	*extra-terrestrial greeter*, *outerspace wellcomist*
Object – Action – Agent	*contactee*, *greeter*

Table 9. Word-formation types vs. possible morphological realizations

In this model, complex units are classified on the grounds of conceptual fields (*Agent*, *Instrument*, *Action*, *Object*, *Locative*, etc.), while the division into word-formation processes is not fundamental. It is precisely in the variety of morphological realizations that the above-explained view of word-formation becomes apparent, in that language users evade productivity constraints to coin the needed NU. In this model, the conceptual labels compete for the naming of an entity, and it is precisely this that justifies the view of word-formation as a creative phenomenon, for it is the speaker that selects the labels and formal composition of NUs (see Table 9). An advantage is that the division into word-formation processes (affixation, compounding, conversion, etc.) is eliminated, which minimizes the heavy weight traditionally given to the formal makeup of coinages. This, plus the sheer weight of conceptual fields and creativity establishes the present model as opposed to other form-oriented approaches.

For the computation of productivity, five onomasiological types are proposed according to the internal constituency of NUs:

- Type I: Complete Complex Structure (hereafter CCS).
- Type II: Incomplete Complex Structure R (hereafter ICSR).
- Type III: Incomplete Complex Structure L (hereafter ICSL).
- Type IV: Simple Structure (hereafter SS).
- Type V: Onomasiological Recategorization (hereafter OR).

Type I is the closest to explicitness of expression (e.g. *landowner*), and occurs in units which encompass, for instance, an Object (*land*), an Action (*own*) and an Agent (*-er*), while type V is the closest to economy of speech, represented by the process of conversion (e.g. *intellectual*^{N} derived from *intellectual*^{AJ}); the remaining three types take intermediate positions. Thus, type II and III occur when the right-hand or the left-hand constituent is left unexpressed, respectively. In the former case, it is the so-called determined constituent of the onomasiological mark that is elided: an item such as *writer* comprises Action (*write*) and Agent (*-er*). In the latter case, i.e. type III, the missing element is the determining constituent of the onomasiological mark: in *honeybee*, only the Object (*honey*) and the Agent (*bee*) are present. As to type IV, it includes items where the onomasiological mark cannot be analysed into the determining and determined parts, and are hence regarded as simple structures in this model, as in *lionhearted*.

These five types are included under a *Word-Formation Type Cluster* (WFTC) which is 100% productive according to the four broad conceptual categories, namely SUBSTANCE, ACTION, QUALITY and CONCOMITANT CIRCUMSTANCE. Since the total productivity of the WFTC is 100%, the internal productivity percentage of each onomasiological type can be measured. This, along with conceptual computations, favours comparisons between different onomasiological types.[14]

Table 10 illustrates the computation of the productivity of the WFTC nouns with label Agent (other labels being Instrument, Quality or Action) in our corpus, where some of the items included under this WFTC are:

– Type I – CCS: *landowner*, *shareholder*, *taxpayer*.

– Type II – ICSR: *committee*, *singer*, *researcher*.

– Type III – ICSL: *headmaster*, *photographer*, *teenager*.

– Type V – OR: *conservative*, *feminist*, *resident*.

Total number of NUs	**106**	**100%**
Type	**NUs**	**Productivity**
Onomasiological Type I – CCS	4	3.773%
1. SUBSTANCE – SUBSTANCE		
(a) Obj – Act – Ag	3	2.830%
2. CONCOMITANT CIRCUMSTANCE – SUBSTANCE		
(a) Temp – Act – Ag	1	0.943%
Onomasiological Type II – ICSR	63	59.433%
1. ACTION – SUBSTANCE		
(a) Act – Ag	62	58.490%

[14] The reader is referred to Štekauer (1998: 180-207; 2001: 10-21, 2005b: 221-22) for a full account of the onomasiological types.

(b) Act – Pattern – Ag	1	0.943%
Onomasiological Type III – ICSL	33	31.132%
1. SUBSTANCE – SUBSTANCE		
(a) Obj – Ag	14	13.207%
2. QUALITY – SUBSTANCE		
(a) Qual – Ag	13	12.264%
3. CONCOMITANT CIRCUMSTANCE–SUBSTANCE		
(a) Loc – Ag	5	4.716%
(b) Temp – Ag	1	0.943%
Onomasiological Type IV – SS	0	0%
Onomasiological Type V – OR	6	5.660%
4. QUALITY – (Ag) – SUBSTANCE	6	5.660%

Table 10. Productivity of the WFTC Agent

As shown, the total 106 items carrying this conceptual label yield the 100% of the productivity. Each unit is encompassed by one of the five onomasiological types and the productivity figures are shown in percentages, so that the global value of different clusters allows comparison within the model. In this case, the most productive type is onomasiological type II, with 59.433% of the productivity, while the least productive one is onomasiological type IV, since it is empty of NUs (i.e. it has 0% productivity). Even if formal comparisons with the rest of the models are unfeasible, it may be helpful to establish rough correspondences where possible. For example, type III can be approximately matched to root/primary compounds, or type V to conversion (*cf.* Štekauer 2001: 12-18). A closer looks reveals that most NUs (62) are included under onomasiological type II with the semantic labels Act – Ag, as in *builder*, *employee* or *survivor*.

5. Discussion

In section 4, six models were described and applied to our corpus for measurement of low productivity. The results disclosed by each of them were presented and described individually with a view to providing an account of areas of low productivity and a general picture of the procedures implemented in each model. The present section discusses the comprehensiveness of the models on the basis of the results provided and procedures used by each of them.

It has been observed, for instance, that the implementation procedures and areas of low productivity largely vary across the models studied, mainly because results are always given in figures which must then be accounted for. Such an interpretation depends on the rationale underlying the model under study which, as seen, may be of various kinds. First, there are models which focus on the frequency of attested items by summing up the total number of derived items (type frequency, see 4.1.1), by summing up the frequencies of such derived items (token frequency, see 4.1.2) or by calculating the average between the base and the derived words (relative frequency, see 4.1.3). The results from our corpus show, among other features, a pre-eminence of noun-forming derivational suffixes for type frequency and of prefixes for token frequency, while affixes forming either nouns or adjectives are prevalent for relative frequency. It may be this discrepancy of results, together with their simplicity of application, that has led most scholars to reject these models for inaccuracy of conclusions (*cf.* Aronoff 1976: 36; Baayen 1992: 110-11; Plag 1999: 11; Bauer 2001: 48-49; Hay and Baayen 2002: 2 *et passim*; Štekauer *et al.* 2005: 12-13; Fernández-Domínguez 2006).

Second, there are models which, together with token frequency counts, add hapax legomena to their computations on the basis that this allows the making of predictions about the productivity of a given affix. Such models may display results in figures (see 4.2.1 and 4.2.2) or on a chart (see 4.2.3), which opens up the opportunity to approach each affix from diverse perspectives. In the case of *P* and *P**, the results obtained are quite parallel, with hardly any inflectional affix and a number of suffixes coming from low-frequency items in both models. P*, alternatively, makes use of graphic representation for display of results which, although advantageous in a sense, prevents further comparison with other models (see Figs. 1 and 2). In consequence, even if probabilistic models stand as a more valid alternative than frequency models, they are inevitably restricted to affixation and are reported to lack a satisfying methodology (see Plag 1999: 33; Bauer 2001: 155; 2005a: 326).

The results of each model explicated, we are now in a position to assess the comparability of results across models. To this end, Table 11 displays all affixes studied, with a dot marking the affixes that are listed by each model among the ten least productive. All in all, 55 affixes have been studied in this paper. Occurrence under one model, which indicates very low correspondence in the results, is the most usual (39 affixes); occurrence under two models, suggesting that models match to a certain extent, is less frequent (14 affixes); while occurrence under three models is rare (2 affixes) and indicates a higher degree of correspondence.

	Frequency			**Hapax**	
	Type	**Token**	**Relat.**	***P***	***P****
AJ-*acy*>N	•				
AJ-*age*>N	•				
AJ-*ancy*>N	•				
N-*ar*>AJ	•		•		
AJ-*ate*>V	•				
cor-N>N				•	•
AJ-*cy*>N	•	•			
dis-V>V				•	
down-N>AV				•	•

N-*ed*>AJ	•				
V-*ee*>N			•		
en-AJ>V				•	•
en-N>V					•
en-V>V					•
AJ-*er*>N	•	•			
AV-*er*>AJ	•				
N-*ery*>N	•				
N-*ern*>AJ				•	
N-*ess*>N	•				
fore-V>V					•
N-*ful*>AJ				•	
N-*ful*>N	•				
AJ-*hood*>N	•				
N-*hood*>N	•				
AJ-*ian*>N	•		•		
AJ-*ic*>AJ			•		
N-*ier*>N	•				
in-N>N		•			
infra-N>N		•			•
N-*ish*>AJ	•		•		
N-*ish*>AJ			•		
V-*ion*>N			•		
AJ-*ism*>N	•	•			
V-*ism*>N	•				
N-*ive*>AJ	•				
D-*ly*>AV	•				
N-*ly*>AJ				•	
V-*ment*>N				•	
off-V>V		•			•
N-*ory*>AJ	•				
V-*our*>N	•		•		
out-N>AJ		•			

N-*ry*>N	•				
N-*ship*>N				•	
AJ-*ster*>N	•				
V-*th*>N	•				
V-*tion*>N	•				
un-AV>AV	•	•			
under-N>PR		•	•		•
up-N>AV	•				
AV-*ward*>AV	•				•
AJ-*wise*>AV				•	
N-*y*>N	•	•	•		
N-*y*>N	•				
V-*y*>V	•				

Table 11. Affixes considered in the study[15]

The results from Table 11 lead to two conclusions. First, except for the matches mentioned above, the differences are significant. In fact, most affixes (39 out of 55) appear under only one model, which means that there is practically no agreement in the results obtained with each of the models. Second, as explained in 4, the models tested usually give an account of how to measure high productivity where, the higher the figure obtained, the more productive the process. No reference is made, however, to whether the opposite correlation also occurs, i.e. whether the lower the figure, the less productive the process. Thus, from the above results, it can be gathered that, while high figures usually equal matching results among models, accuracy decreases parallel to the figures.

In contrast with the frequency or probabilistic models, a different trend is represented by the onomasiological approach (see 4.3), where the focal point is meaning. Here, all traditional word-formation processes are arranged into five onomasiological types not aimed at comparison with traditional productivity measurements. Instead, productivity is computed on the basis of the word's semantic, and not structural, makeup. In our case, the onomasiological model was applied to all nouns in our corpus carrying the semantic label Agent. Table 12 illustrates that type IV is noticeably the least productive one (0%) followed by type I (3.773%) which, as stated above, would indicate a low productivity of the traditional process of conversion (type V). As observed above, an analysis in terms of conversion in the rest of the models would be inconceivable due to the lack of formal marks.

[15] Please note that this table includes only results from compatible models, which excludes the onomasiological model (see Table 10) and global productivity (see Figs. 1 and 2). Affixes occurring more than once are shaded grey.

Type	Productivity
Onomasiological Type I – CCS	3.773%
Onomasiological Type II – ICSR	59.433%
Onomasiological Type III – ICSL	31.132%
Onomasiological Type IV – SS	0%
Onomasiological Type V – OR	5.660%

Table 12. Summary of results from the onomasiological model

In view of the characteristics of application of the existent models, a number of limitations become apparent in relation to their comprehensiveness for low productivity measurement. First, most models seem to examine profitability and fall short for the study of availability (see fn. 5). Although not explicitly specified in the literature, this can be seen in that most models are based on formulas for computing the number of items created by a given process (profitability), while the issue of whether the process studied is currently in use (availability) is neglected by this methodology. Both notions are equally relevant in productivity and, as suggested in the literature, they are issues which require a different treatment: the fact that a process has proved highly productive does not necessarily mean that it should be available at present (Bauer 1983: 55; Corbin 1987: 177; Plag 2003: 52). Despite being a complex notion, this is an area still to be explored, and one without which the study of productivity is incomplete.

A second limitation, common to all models except for the onomasiological, is that they seem to be designed for the study of affixation alone. In some cases, the models can be applied to items obtained from other word-formation processes without difficulty, since the procedure is easy to apply. For instance, in type frequency, and similarly in token frequency, computations of the productivity for compounding are made by summing up the number of types and tokens. Nevertheless, in other cases, application of the formulas to items from other word-formation processes where there is no detachable element comparable to an affix poses difficulties (Fernández-Domínguez 2006). This can be illustrated by trying to obtain the productivity for compounding or acronymy in relative frequency, where the frequencies of the derived unit and of the base unit are needed. Provided the frequency of the derived unit is available, the main problem remains unsolved: how is the frequency of the base obtained? The problem holds in Baayen's models: if we apply *P* to the suffix -*ment*, retrieval of the hapaxes for this suffix would require searching for all words ending in -*ment*, checking them individually and summing the partial figures. This cannot be done with processes such as conversion, where there is no detachable element that can be searched for and finally summed up (but see type V of the onomasiological approach). This shortcoming seems to hold in all processes except for affixation.

As seen in 4.3, however, this is not the case with the onomasiological approach, since its computations are based on semantics. While affixes are exclusive to affixation, or multiple words to compounding or blending, semantic labels occur in all word-formation processes in one form or another. Štekauer's approach can thus account for all word-formation processes, taking productivity as a non-universal term, as an epiphenomenon.[16] Indeed, the productivity of a process is agreed to be influenced by multiple factors, such as constraints or the speech community, which may affect its output, but these are rarely taken into consideration by other models. Therefore, it seems to us that such a semantic-based approach overcomes the limitations of strictly form-based models, which often seem to lack objectivity due to the self-imposed requirement to choose among formal processes (*cf.* Štekauer 2001: 29-30).

6. Conclusions

The tendency among scholars to focus on high productivity measurement has made of low productivity a somewhat neglected area. In this context, this paper has served to show that this is a developing area attracting the attention of scholars where agreement is still to be reached. From this study the following conclusions can be gathered:

i.) Each model is based on different computations and each of them provides different areas of low productivity, which leaves an incongruent picture of the phenomenon. This, however, should not be perceived as an inherent drawback to the models, but rather as an evidence of the number of proposals that are being made in this area, which only serves to show the interest that productivity measurement is generating nowadays.

ii.) In all models, the interpretation of figures for low productivity resulting from their application is ambiguous. Whereas high figures unequivocally correspond to high productivity, it is not entirely clear whether low figures correspondingly match low productivity or whether they imply a decrease in measurement accuracy.

iii.) Every model has advantages but also disadvantages in its application. For example, conventional models are restricted to affixation measurement, which overlooks many other means of linguistic innovation and thus gives incomplete results for productivity. Similarly, the probabilistic models seem to rely too much on figures for interpretation of data. On the other hand, the semantic-based approach of the onomasiological theory is at the same time a plus and a minus; a plus in that it represents a step forward in productivity measurement, and a minus in that it isolates such an approach from the rest.

iv.) Little attention is paid to non-quantitative aspects in these computations, such as availability or the needs of the speech community. In our opinion, it is necessary for models to come to take these notions into consideration for a full understanding of morphological productivity.

---

[16] Plag (1999: 11-13) uses the term *epiphenomenon* to emphasize that productivity is influenced by various factors and is not an absolute term.

Works Cited

Aronoff, Mark 1976: *Word Formation in Generative Grammar. Linguistic Inquiry Monograph 1.* Cambridge: MIT Press.

——— 2001: 'Morphology between lexicon and grammar'. Geert Booij, Christian Lehmann and Joachim Mudgan, eds. *Morphologie. Morphology. Ein Internationales Handbuch zur Flexion und Wortbildung. An International Handbook on Inflection and Word-Formation.* Vol. 1. Berlin and New York: Walter de Gruyter. 344-49.

Aronoff, Mark and Frank Anshen 2001: 'Morphology and the lexicon: lexicalization and productivity'. Andrew Spencer and Arnold M. Zwicky, eds. *The Handbook of Morphology*. Oxford: Blackwell Publishers. 237-47.

Aronoff, Mark and Kirsten Fudeman 2005: *What is Morphology?* Oxford: Blackwell Publishers.

Baayen, Harald 1992: 'Quantitative aspects of morphological productivity'. Geert Booij and Jaap van Marle, eds. *Yearbook of Morphology 1991.* Dordrecht: Kluwer. 109-49.

——— 1993: 'On frequency, transparency and productivity'. Geert Booij and Jaap van Marle, eds. *Yearbook of Morphology 1992.* Dordrecht: Kluwer. 181-208.

——— 1994: 'Productivity in language production'. *Language and Cognitive Processes* 9.3: 447-69.

——— 2003: 'Probabilistic approaches to morphology'. Rens Bod, Jennifer Hay, and Stefanie Jannedy, eds. *Probabilistic Linguistics*. Cambridge: MIT Press. 229-87.

Baayen, Harald and Rochelle Lieber 1991: 'Productivity and English derivation: a corpus-based study'. *Linguistics* 29: 801-44.

Bauer, Laurie 1983: *English Word-Formation*. Cambridge: Cambridge UP.

——— 2001: *Morphological Productivity*. Cambridge: Cambridge UP.

——— 2003 (1988): *Introducing Linguistic Morphology*. Edinburgh: Edinburgh UP.

——— 2005a: 'Productivity: theories'. Pavol Štekauer and Rochelle Lieber, eds. *Handbook of Word-Formation*. Dordrecht: Springer. 315-34.

——— 2005b: 'The borderline between derivation and compounding'. Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer and Franz Rainer, eds. *Morphology and its Demarcations*. Amsterdam and Philadelphia: John Benjamins. 97-108.

Burchfield, Robert W., ed. 1987: *The Compact Edition of the Oxford English Dictionary. A Supplement to the Oxford English Dictionary*. Oxford: Clarendon Press.

Chomsky, Noam 1970: 'Remarks on nominalization'. Roderick A. Jacobs and Peter S. Rosenbaum, eds. *Readings in English Transformational Grammar*. Waltham: Ginn. 184-21.

Corbin, Danielle 1987: *Morphologie Dérivationelle et Structuration du Lexique*. 2 vols. Tübingen: Niemeyer.

Fernández-Domínguez, Jesús 2006: *A Corpus-Based Study of Morphological Productivity Measurements in Present-Day English Word-Formation*. Unpublished M.A. Thesis. University of Jaén, Spain.

Giegerich, Heinz J. 1999: *Lexical Strata in English. Morphological Causes, Phonological Effects*. Cambridge: Cambridge UP.

Hay, Jennifer and Harald Baayen 2002: 'Parsing and productivity'. Geert Booij and Jaap van Marle, eds. *Yearbook of Morphology 2001*. Dordrecht: Kluwer. 203-35.

——— 2003: 'Phonotactics, parsing and productivity'. *Revista di Linguistica* 15.1: 99-30.

Hohenhaus, Peter 2005: 'Lexicalization and institutionalization'. Pavol Štekauer and Rochelle Lieber, eds. *Handbook of Word-Formation*. Dordrecht: Springer. 353-73.

Jensen, John T. 1990: *Morphology. Word Structure in Generative Grammar*. Amsterdam and Philadelphia: John Benjamins.

Kastovsky, Dieter 1986: 'The problem of productivity in word-formation'. *Linguistics* 24: 585-600.

——— 2005: 'Hans Marchand and the Marchandeans'. Pavol Štekauer and Rochelle Lieber, eds. *Handbook of Word-Formation*. Dordrecht: Springer. 99-124.

Katamba, Francis 1993: *Morphology*. London: MacMillan Press Ltd.

Kilgarriff, Adam 1996: 'BNC frequency list'. <http://www.kilgarriff.co.uk/BNC_lists/all.num.gz> (Accessed 6 April, 2006).

Kiparsky, Paul 1983: 'Word formation and the lexicon'. Frances Ingemann, ed. *Proceedings of the 1982 Mid-America Linguistics Conference*. Lawrence: Kansas UP. 3-29.

Leech, Geoffrey, Paul Rayson and Andrew Wilson 2004: 'Rank frequency lists of words within word classes (parts of speech) in the whole corpus'. <http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html> (Accessed 2 February, 2006).

Lieber, Rochelle 2004: *Morphology and Lexical Semantics*. Cambridge: Cambridge UP.

Lipka, Leonhard 1977: 'Lexikalisierung, Idiomatisierung und Hypostasierung als Probleme einer synchronischen Wortbildungslehre'. Herbert E. Brekle and Dieter Kastovsky, eds. *Perspektiven der Wortbildungsforschung*. Bonn: Bouvier. 155-64.

Lyons, John 1977: *Semantics*. 2 vols. Cambridge: Cambridge UP.

Marchand, Hans 1969 (1960): *The Categories and Types of Present-Day English Word-Formation*. München: Beck.

Murray, James, Henry Bradley, William A. Craigie and Charles T. Onions, eds. 1971: *The Compact Edition of the Oxford English Dictionary*. 2 vols. Oxford: Clarendon Press.

Plag, Ingo 1999: *Morphological Productivity: Structural Constraints on English Derivation*. Berlin and New York: Mouton de Gruyter.

——— 2003: *Word-Formation in English*. Cambridge: Cambridge UP.

——— 2006: 'Productivity'. Keith Brown, ed. *Encyclopedia of Language and Linguistics*. Boston: Elsevier. 121-28.

Plag, Ingo, Christiane Dalton and Harald Baayen 1999: 'Morphological productivity across speech and writing'. *English Language and Linguistics* 3.2: 209-28.

Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech and Jan Svartvik 1985: *A Comprehensive Grammar of the English Language*. New York: Longman.

Rainer, Franz 2001: 'Produktivitätsbeschränkungen'. Geert Booij, Christian Lehmann and Joachim Mudgan, eds. *Morphologie. Morphology. Ein Internationales Handbuch zur Flexion und Wortbildung. An International Handbook on Inflection and Word-Formation*. Vol. 1. Berlin and New York: Walter de Gruyter. 877-85.

——— 2005a: 'Constraints on productivity'. Pavol Štekauer and Rochelle Lieber, eds. *Handbook of Word-Formation*. Dordrecht: Springer. 335-52.

——— 2005b: 'Semantic change in word formation'. *Linguistics* 43.2: 415-41.

Roeper, Tom 2005: 'Chomsky's *Remarks* and the transformationalist hypothesis'. Pavol Štekauer and Rochelle Lieber, eds. *Handbook of Word-Formation*. Dordrecht: Springer. 125-46.

Rose, James H. 1973: 'Principled limitations on productivity in denominal verbs'. *Foundations of Language* 10: 509-526.

Scalise, Sergio 1984: *Generative Morphology*. Dordrecht: Foris.

Schultink, Henk 1961: 'Produktiviteit als morfologisch fenomeen'. *Forum der Letteren* 2: 110-125.

Simpson, John A. and Edmund. S. C. Weiner, eds. 1989: *The Oxford English Dictionary. Second edition*. Oxford: Clarendon Press.

Spencer, Andrew 1991: *Morphological Theory*. Oxford: Blackwell Publishers.

Štekauer, Pavol 1998: *An Onomasiological Theory of English Word-Formation*. Amsterdam and Philadelphia: John Benjamins.

——— 2001: 'Fundamental principles of an onomasiological theory of English word-formation'. *Onomasiology Online* 2: 1-42. <http://www.ku-eichstaett.de/SLF/EngluVglSW/stekauer1011.pdf > (Accessed 28 July, 2005).

——— 2005a: *Meaning Predictability in Word-formation: Novel, Context-Free Naming Units*. Amsterdam and Philadelphia: John Benjamins.

——— 2005b: 'Onomasiological approach to word-formation'. Pavol Štekauer and Rochelle Lieber, eds. *Handbook of Word-Formation*. Dordrecht: Springer. 207-32.

——— 2006: 'Onomasiological theory of word-formation'. Keith Brown, ed. *Encyclopedia of Language and Linguistics*. Boston: Elsevier. 34-7.

Štekauer, Pavol, Don Chapman, Slávka Tomaščíková and Štefan Franko 2005: 'Word-formation as creativity within productivity constraints: sociolinguistic evidence'. *Onomasiology Online* 6: 1-55. <http://www1.ku-eichstaett.de/SLF/EngluVglSW/stekauer1051.pdf> (Accessed 28 December, 2006).

Tournier, Jean 1985: *Introduction Descriptive à la Lexicogénétique de L'Anglais Contemporain*. Genève: Editions Slatkine.

Yip, Moira 1998: 'Identity avoidance in phonology and morphology'. Stephen G. Lapointe, Diane K. Brentari and Patrick M. Farrell, eds. *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI Publications. 216-46.

Received 10 October 2006
Revised version received 12 March 2007