

Quantitative aspects of morphological productivity*

HARALD BAAYEN

1. INTRODUCTION

Research into the phenomenon of morphological productivity, "the possibility for language users to coin, unintentionally, a number of formations which are in principle uncountable" (Schultink 1961), has mainly focused on the qualitative factors which jointly determine the productivity of word formation rules. It is well known that word formation processes are subject to various syntagmatic conditions. Booij (1977) develops a typology of such conditioning factors, distinguishing between rule-specific and rule-independent restrictions on the one hand, and between restrictions pertaining to phonological, stratal and syntactic characteristics on the other.¹ The rôle of paradigmatic factors is discussed in van Marle (1985). He points out that (roughly) synonymous affixes tend to select their base words from complementary domains. Hence they can be analyzed as mutually affecting their respective degrees of productivity.

Other kinds of conditioning factors involved are semantic coherence and contextual appropriateness. The importance of semantic coherence is stressed by Aronoff (1976). He shows that there is a direct link between semantic coherence and productivity. The words generated by the more productive rules are semantically highly predictable, formations covered by the less productive and unproductive rules are often characterized by various unpredictable readings. When complex words assume such diverse meanings that the core meaning of the morphological category becomes opaque, this may cause speakers to become uncertain as to the semantic function of the corresponding word formation rule, with the effect that they are less likely to use it. Van Marle (1988) argues that the Dutch suffix *-lijk*, as in *waarlijk* 'truly', has lost its productivity in precisely this way.

The rôle of contextual appropriateness is explicitly taken into account in the version of Coseriu's (1970, 1975) theory of 'System, Norm und Rede' developed by Burgschmidt (1977). Burgschmidt discusses the phenomenon that the extent of use of well-formed complex items is a function of the social context. For instance, while *rentenempfangsberechtigt* 'pensionable' is acceptable in the context of official language, the use of *suppenempfangsberechtigt* 'entitled to receiving soup' in the family circle is ridiculous. With respect to derivation, we may mention the Dutch suffix *-erd*, which is used to coin slightly pejorative personal names in Dutch such as *bangerd* and *dikkerd* from *bang* 'afraid' and *dik* 'fat', respectively. Interestingly, this suffix shows up with only seven types in the written language of the Eindhoven corpus (henceforth EC), a corpus of some 600 000 word forms. Even though *-erd* is judged to be productive (see e.g. Schultink 1962: 200–205), it is not exten-

sively used in written language, which usually requires a more formal style in which such formations are inappropriate. In other words, for a word formation rule to be fully productive, it should be appropriate in a sufficiently broad range of styles. In fact, speech style is but one of a number of non-linguistic factors which may codetermine the productivity of word formation rules, such as the socio-economic status of the language user, his or her attitude towards the morphological processes of the language,² and the pragmatic factor of the usefulness of the concepts associated with the complex words generated by a given rule. For instance, van Santen and de Vries (1981) argue that the absence in dictionaries of Dutch of many formations in the productive suffix *-ster*, which forms female personal nouns, is due to the low pragmatic usefulness of such female personal names.³ In what follows, I will use the expression 'extent of use' to refer to the combined effects of these various non-linguistic factors on the 'global productivity' of word formation rules, that is, the overall productivity as the outcome of the interaction of linguistic and non-linguistic factors.

The notion of morphological productivity has received considerable clarification from the study of the various kinds of restrictions which have been found to condition word formation rules. In a qualitative sense, the productivity of a word formation rule can be said to be inversely proportional to the number of conditioning factors in force (Booij 1977). Nevertheless, the quantitative outcome of the interaction of the — often highly heterogeneous — conditioning factors has remained rather obscure. The aim of the present paper is to clarify some of the issues involved in the quantification of morphological productivity. In Section 2 three complementary measures of morphological productivity are developed, which make use of the statistical information contained in the empirical frequency distributions of morphological categories in text corpora. Section 3 relates our findings to the theory of the mental lexicon, and Section 4 reviews three models in which the relevance of token frequencies is recognized, the models developed by Anshen and Aronoff (1988), Bybee (1985, 1988) and Rumelhart and McClelland (1986).

2. THE QUANTITATIVE ANALYSIS OF MORPHOLOGICAL PRODUCTIVITY

Any measure of morphological productivity that is of linguistic interest will have to satisfy a number of requirements. First, such a measure should provide a ranking of word formation processes that is in general correspondence with a ranking based on linguistic intuitions. For instance, a measure that ranks the degree of productivity of English *-ity* above that of English *-ness* is clearly unsatisfactory. Secondly, such a measure should express "the statistically determinable readiness with which an element enters into new combinations." (Bolinger 1948: 18). Third, taking into account those formations which are characterized by formally or semantically idiosyncratic properties should have the effect of lowering the value of the productivity measure. And fourth, such a measure should shed light on the

empirical fact that productivity cannot be simply measured in terms of type frequencies. Consider, for instance, the Dutch suffixes *-sel*, judged to be productive by Geerts *et al.* (1984: 93), and *-te*, judged to be unproductive by Schultink (1962), which are found with roughly the same number of types, 44 and 39 respectively, in the EC. Even more striking is the fact that action nouns with vocalic alternation, such as *spel* from *spe(e)l-en* 'to play', are represented by some 100 types, that is roughly 2.3 times the number of types in productive *-sel*, even though vocalic alternation is an unproductive process in modern standard Dutch.

In order to come to grips with the quantitative aspects of productivity, an analysis of the word frequency distributions of morphological classes is required. This implies that, in addition to a simple count of the number of different formations with a given affix, we also have to take the frequencies of use of these formations into account. Some authors, for instance Schultink (1961) and Rainer (1988) have argued that token frequencies are irrelevant to the problem at hand. Others, notably Harwood and Wright (1956), Bradley (1979), Bybee (1985) and Anshen and Aronoff (1988) have sought to relate productivity and token frequency. Following their lead, we will subject the type and token frequencies of the formations in a given affix to a principled statistical analysis, and show that productivity and frequency are indeed closely correlated.

In what follows, we will make use of two corpora, the Dutch Eindhoven corpus (EC), and the English Cobuild corpus (CC). The EC is a corpus of some 600 000 word forms of written language. It covers text fragments taken from daily and weekly newspapers, from magazines, popular scientific prose and novels (see Uit den Boogaart 1975). The CC, with 18 000 000 word forms, is taken from both spoken and written language (25% spoken, 75% written), and contains predominantly British English. It covers "broadly general, rather than technical, language, current usage, from 1960, and preferably very recent; 'naturally occurring' text, not drama; prose, including fiction and excluding poetry; adult language, 16 years and over;" (Renouf 1987: 2). The use of corpora is motivated by the fact that they offer information about the token frequencies of the types, and by the fact that they are more trustworthy than dictionaries with respect to the words in current use. On the one hand, corpora contain words of the sort that dictionaries typically do not list, notably words formed with highly productive affixes. On the other hand, as pointed out by Anshen and Aronoff (1988: 645), dictionaries may list words which are not used in actual speech. Even though Walker (1936) lists 23 words in *-ivity* and 27 words in *-ibleness*, only the words in *-ivity* are attested in the Kučera and Francis (1967) corpus.

The first step towards a quantitative analysis⁴ of productivity is to select from some fixed corpus all occurrences (tokens) of the formations (types) with the morphological constituency of interest. Let V denote the number of such types, and N the associated number of tokens. The V types in a sample can be ranked according to decreasing token frequency f_i , such that $f_i \geq f_{i+1}$ for all i ($i = 1, 2, \dots, V$). Types with the same token frequency are ordered

arbitrarily. For instance, the frequency distribution of simplex nouns in EC could be summarised as shown here for the first ten types.

<i>jaar</i>	$f_1 = 1237$	'year'	<i>dag</i>	$f_6 = 577$	'day'
<i>mens</i>	$f_2 = 923$	'human'	<i>plaats</i>	$f_7 = 483$	'place'
<i>man</i>	$f_3 = 743$	'man'	<i>hand</i>	$f_8 = 472$	'hand'
<i>tijd</i>	$f_4 = 671$	'time'	<i>land</i>	$f_9 = 464$	'land'
<i>vrouw</i>	$f_5 = 603$	'woman'	<i>heer</i>	$f_{10} = 413$	'lord'

A more concise form for summarizing the data is to group the f_i such that all n_r types for which $f_i = r$ are brought together in a frequency class r . The frequency classes are then listed according to increasing rank r , yielding a so-called grouped frequency distribution. The grouped frequency distribution of words with the Dutch suffix *-heid*, which forms abstract nouns from adjectives (e.g. *snelheid* 'speed', from *snel* 'quick'), as found in the EC, has been listed in Table 1. The general shape of this grouped frequency distribution is not unfamiliar from literary studies on texts as a whole (see e.g. Herdan 1964). Note that this distribution is highly skewed to the lower ranks r : the value of n_r decreases for increasing r , rapidly for the lower values of r , slowly for the higher ranks.

The grouped frequency distribution is a rich source of information. To begin with, the number of tokens N and the total number of types V in the sample is obtained from the grouped frequency as follows:

$$(1) \quad N = \sum_r r n_r,$$

$$(2) \quad V = \sum_r n_r.$$

Applied to the above distribution of abstract nouns in *-heid*, the 466 different types are obtained by summation of the entries in the columns of Table 1 labelled n_r . By first calculating, for each frequency r , the number of tokens m_r that the n_r types with this token frequency r contribute to the overall distribution, followed by summation for all frequencies r over the products m_r , the total number of tokens N is obtained.

At this point we should pause to note that the way in which we have

Table 1. Grouped frequency distribution of *-heid* (EC)

r	n_r	r	n_r	r	n_r	r	n_r	r	n_r	r	n_r
1	256	7	9	15	2	27	1	47	2	85	1
2	71	8	7	16	1	28	1	49	1	106	1
3	37	9	2	17	2	29	2	54	1	114	1
4	24	10	5	21	2	34	1	69	1	163	1
5	12	12	1	23	1	37	1	70	1	—	—
6	11	13	3	26	1	43	1	75	1	—	—

obtained our data involves two sampling stages. In the first stage, some corpus is selected, a corpus which, ideally, is a representative sample of the language under investigation. In the second stage, a subset of tokens is extracted from this corpus, namely, all word forms (tokens) with some particular affix. We will refer to the corpus as the frame sample, and to the extracted set of tokens as the item sample. It is important to realize that the values of N and V , as calculated from the item sample, depend on the size of the frame sample. For larger frame samples, larger values of N and V are to be expected for the item sample. Consequently, for some fixed morphological process, V can be viewed as a function of N : for increasing numbers of tokens in the item sample, obtained by increasing the frame sample, V will also increase.

In the light of the fact that V is a function of N , and writing $V^{(N)}$ to emphasize this fact, the mathematical characterization of this function is of interest. Figure 1 shows that $V^{(N)}$ is a non-linear function of N . Attempts to express V as some simple function of N , for instance, Herdan's (1964: 145–147) 'law' $V = N^\beta$, fail, especially for large values of N (for a detailed discussion see Baayen 1989). This is unfortunate, since such a function would yield the means to obtain two important characteristics of item samples, namely (i) an estimate of the growth rate of V at any point N , and (ii) an estimate of the number of types S in the population being sampled. Both this estimate of the growth rate of V and the estimate of S are relevant with respect to the quantitative analysis of productivity. The growth rate is a measure of the likelihood of coming across new types, and hence a promising

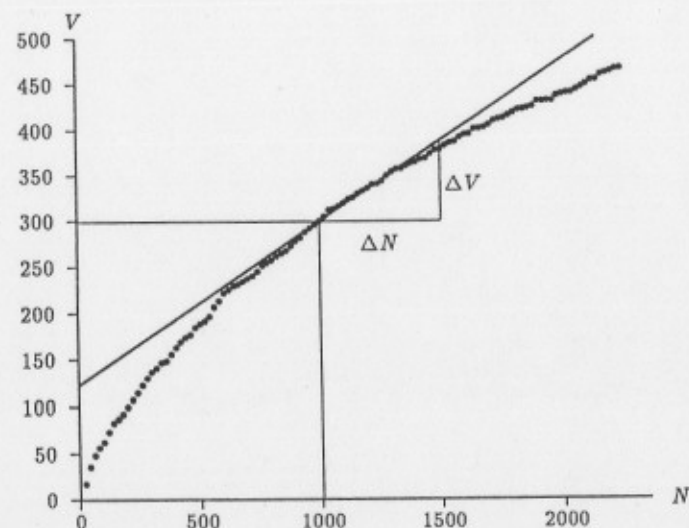


Figure 1. The growth curve of *-heid* in the EC ($N = 2251$, $V = 466$). The growth rate of V for sample size 1000 can be expressed in terms of the slope $\Delta V/\Delta N = 0.177$ of the tangent to the curve in the point (1000, 299).

statistical measure of the degree of productivity. An estimate of S allows us to compare the number of attested types with the number of possible types, a comparison which sheds light on the extent to which the number of actual words exhausts the number of possible words.

In principle, the growth rate of V at a particular sample size M can be obtained by calculating the derivative of $V^{(N)}$ in the point $(M, V^{(M)})$, that is, the slope of the tangent to the growth curve of V at sample size M . Figure 1 illustrates this idea. On the horizontal axis we have the size of the item sample N , the number of tokens with — in this case — the suffix *-heid* in the EC. On the vertical axis the number of different types V is plotted. The number of types V increases with N , quickly for the lower sample sizes, slowly for the larger ones. For $N = 1000$ we count 299 different types. The tangent to the growth curve of V in the point $(1000, 299)$ characterizes at what rate new types are being encountered during the sampling process at the time that N equals 1000. For smaller values of N , the steepness of the tangent increases, for larger samples it decreases. This corresponds with the fact that in the initial sampling stages most tokens entering the sample represent new types, while as the sampling process continues the probability that a given type has not been encountered before steadily decreases. Since the steepness of the tangent is expressed by its slope, the growth rate of V can be formalized as $\Delta V/\Delta N$.

The population number of types S is the second statistic of interest. Mathematically, S is estimated by considering the limit of $V^{(N)}$ for $N \rightarrow \infty$, that is, by calculating the number of types that in theory are expected to occur in a sample of unlimited size. There are two possible outcomes here. The growth curve may flatten out, implying that the number of types in the population, the size of the set of morphologically possible types, is finite. Of course, the limiting value of V may be quite small, indicating that an unproductive morphological category is sampled, or it may be very large, a state of affairs that would indicate that a productive process is being analysed. The other possibility is that the growth curve does not have a finite limit. In that case we are dealing with a very productive process that theoretically gives rise to an infinite number of possible types.

What we need, then, in order to obtain reliable estimates of both the growth rate of V for a given sample size, and of the limiting value of V , is a mathematical description of the growth curve itself. In Section 2.1 a technique is introduced by means of which an estimate of the growth rate of V can be obtained. It will be shown that this growth rate does extremely well as a measure of the degree of productivity. In Section 2.2. a statistical theory based on Zipf's law is discussed by means of which estimates of the possible number of types S can be calculated. The results obtained shed new light on the relevance of the notion 'number of possible types' for the theory of productivity. In Section 2.3. we turn to the interpretation of V , and introduce the notion of 'global productivity'.

2.1. The growth rate \mathcal{P} of the vocabulary V

The growth rate of V for a particular sample size M can be obtained by differentiating the function $V^{(N)}$ in the point $(M, V^{(M)})$. Unfortunately, there is no simple formula that expresses $V(N)$ in terms of N . However, it can be shown (Kalinin 1965, Good and Toulmin 1956, Efron and Thisted 1976) that $V^{(N)}$, the number of types for arbitrary sample size N , can be expressed as a function f of $V^{(M)}$ and $n_r^{(M)}$, $r = 1, 2, \dots$, for some fixed value of M for which these statistics are available:

$$(3) \quad V^{(N)} = f(V^{(M)}, n_1^{(M)}, n_2^{(M)}, \dots)$$

In other words, when we know the size of a given item sample, M , and given the number of different types V and the numbers of types that occur once, twice, etc. in that sample, we can in principle calculate the number of types expected to be counted in other item samples of size N , $N \neq M$ given the function f . Note that essential use is being made of the way in which the N tokens of the item sample are distributed over the various types: rather than expressing $V^{(N)}$ directly in terms of N , we are making use of the extra information contained in the grouped frequency distribution.

Given the function f , the growth rate of the vocabulary can be obtained by differentiation. Interestingly, the resulting growth rate $\mathcal{P} = f'$ for sample size N is estimated by the simple expression

$$(4) \quad \mathcal{P} = n_1/N,$$

where n_1 is the number of types occurring only once in the item sample of N tokens, the so-called hapaxes. Returning to Figure 1, this theory implies that

$$\Delta V/\Delta N = n_1/N.$$

In fact, the slope of the tangent to the growth curve in $(1000, 299)$ in Figure 1, 0.177, was calculated on the basis of (4). That a reasonable tangent to an empirical growth curve is obtained suggests that the theoretical model underlying the derivation of (4) is sensitive enough for the present purpose. An important property of \mathcal{P} is that it expresses in a very real sense the probability that new types will be encountered when the item sample is increased. Hence it is not simply a summary statistic like e.g. the mean token frequency that characterizes the central tendency of the distribution. The main interest of \mathcal{P} is that it is the quantitative formalization of the linguistic notion of morphological productivity. As such it satisfies our second criterion for a sensible quantitative measure of productivity, namely that it express the statistical readiness with which new formations are encountered.

We can test whether \mathcal{P} provides a measure that ranks affixes according to their degree of productivity in a way that accords with linguistic intuitions by applying it to sets of rival affixes. The choice for rival affixes is motivated by the fact that for such affixes, which attach to and form words of the same category, and which have more or less the same semantic contribution, effects on the degree of productivity arising from differences in the extent of use or from differences in word category are largely eliminated. Hence the

differences in productivity being measured are most closely related to the linguistic factors determining the qualitative productivity of the rules involved. Table 2 lists the number of tokens N , the number of types V , the number of hapaxes n_1 and the growth rate \mathcal{P} for the Dutch rival affixes *-te* (*warmte* 'warmth') and *-heid* (*snelheid* 'speed'), and the English suffixes *-ness* (*happiness*) and *-ity* (*purity*). For both languages the data on the category of simplex nouns have been added, since these categories are unproductive by definition: there are no morphological word formation processes by means of which the sets of simplex items can be extended. Hence the growth rates of such sets provide a means for weighting the growth rates of the sets of morphologically complex formations. For complex words using a productive word formation process, a growth rate that is significantly larger than that of the simplex items of the same word category is expected. For unproductive formations, the growth rate should not significantly exceed that of the corresponding simplex class. These predictions are born out by Table 2. We find that the productive suffixes *-heid* and *-ness* show up with the highest growth rates. For these suffixes, the probability that new formations will be encountered when the item sample is extended is largest.⁵ In contrast, unproductive *-te* has a growth rate that is only marginally larger than that of Dutch simplex nouns, as shown by the \mathcal{P} values in Table 2.⁶ This is in accordance with the general assessment that this suffix is not, or only marginally productive. Turning to the English data, we find that the growth rate of *-ity*, 0.0007, is roughly 1/6 of that of *-ness* (0.0044). This is in accordance with the fact that *-ity* is less productive than *-ness* (Aronoff 1976). On the other hand, the growth rate of *-ity* is far larger than that of the English simplex nouns than in the case of Dutch *-te* (by a factor 7.1 for *-ity*, 1.6 for *-te*). This is in line with the fact that, although *-ity* is less productive than *-ness*, it is not unproductive, especially when it attaches to adjectives in *-ic*, *-al*, *-able/-ible*.⁷ These comparisons illustrate that \mathcal{P} provides a correct

Table 2. \mathcal{P} for de-adjectival abstract nouns in English and Dutch. Note the difference in the frame sample size F , 18 000 000 for English, 600 000 for Dutch

affix	N	V	n_1	\mathcal{P}	\mathcal{P} ratios	productive
English ($F = 18\,000\,000$)						
simplex nouns	2 142 828	5543	128	0.0001	1.0	-
<i>-ity</i>	42 252	405	29	0.0007	7.0	±
<i>-ness</i>	17 481	497	77	0.0044	44.0	+
Dutch ($F = 600\,000$)						
simplex nouns	37 836	1495	294	0.008	1.0	-
<i>-te</i>	758	39	10	0.013	1.6	-
<i>-heid</i>	2 251	466	256	0.114	14.2	+

ranking of affixes according to their degree of productivity, thus satisfying our first criterion for a quantitative measure of productivity.

The third and fourth criteria are also met. First consider the effect of the presence of idiosyncratic formations on the value of \mathcal{P} . Since such formations typically have high token frequencies — a well-known fact discussed in Aronoff (1976, 1982) — their presence raises N , without influencing n_1 . Hence, such formations have the effect of lowering the value of $\mathcal{P} = n_1/N$, giving expression to the fact that the presence of idiosyncratic formations is detrimental to the degree of productivity of word formation rules.

With respect to the fourth criterion, which requires an explanation for the fact that intuitions concerning productivity cannot be directly linked with the numbers of types V , we may note the following. The number of types counted for some sample size N does not tell us anything about the rate at which new types will appear in larger samples. For some affixes, this rate may be minimal, for others \mathcal{P} may have a substantial value, depending on the characteristics of the underlying population. When we compare the Dutch derivational categories of patient nouns in *-sel* (44 types), de-adjectival abstract nouns in *-te* (39 types), and action nouns with vocalic alternation (100 types), it is the value of \mathcal{P} , and not that of V , which tells us that only *-sel* is productive (Table 3). The measure \mathcal{P} provides the means for distinguishing between productive and unproductive affixes, irrespective of the number of types.

The productivity measure \mathcal{P} has one disadvantage, however. Since $\mathcal{P} = n_1/N$ is itself a function of N , its value depends on the size of the item sample. Since the growth curve of $V^{(N)}$ flattens out for increasing N , we know that \mathcal{P} will become smaller for increasing N . For instance, when half the tokens in the item sample of *-heid* have been counted, a value of $\mathcal{P} = 0.170$ is obtained. When all the tokens of the item sample are taken into account, the value of \mathcal{P} drops to 0.114. Interestingly, the rank r with the highest number of types n_r , the mode of the grouped frequency distribution, lies at 1 for both sampling moments. This is characteristic of productive affixes. In contrast, unproductive processes evidence a shift in the value of the mode. For small samples, the mode may equal unity, but for larger values of N the mode assumes larger values. This is illustrated in Figure 2, which shows the histograms of the grouped frequency distribution of the action nouns with vocalic alternation in the EC. When roughly half the tokens in the item

Table 3. Growth rate \mathcal{P} and number of types V for four frequency distribution

	V	\mathcal{P}
action nouns with vocalic alternation	100	0.004
simplex nouns	1495	0.008
<i>-te</i>	39	0.013
<i>-sel</i>	44	0.080

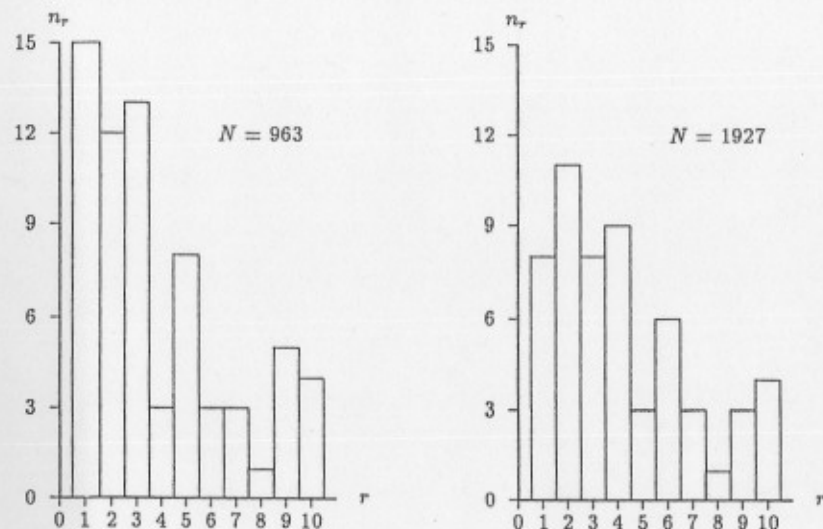


Figure 2. Grouped frequency distributions for nouns with vocalic alternation in the Dutch Eindhoven corpus, at $N = 963$ and $N = 1927$. The histograms, which show the first 10 ranks only, illustrate how the mode of the distribution of this unproductive category shifts to the right for increasing N .

sample have been processed, the number of types with token frequency 1 outranks the number of types with any other token frequency r . However, when all 1927 tokens have been sampled, we find that the mode has shifted to $r = 2$. This implies that the greater part of the types in the population has been sampled. For still larger values of N the situation will eventually occur in which all types of the (finite) population will have been sampled at least twice, a situation for which $n_1 = 0$ and for which the mode will have shifted still further to some higher rank r . Consequently, we may expect that for large N the value of our measure of productivity, \mathcal{P} , which is already smaller than that of the set of simplex nouns in Dutch (0.004 versus 0.008), will become zero for this unproductive process.

We have seen that the mode of the grouped frequency distribution of unproductive affixes shifts to the right for large N , causing \mathcal{P} to become 0, a fact which is a natural consequence of the finite character of the number of types in the underlying population. Since, at least in the calculus of linguistic theory, productive word formation rules give rise to an infinite number of types, it is reasonable to assume that for increasing N new types will continue to be sampled again and again. Consequently, n_1 is expected to retain a value greater than zero even for very large N . This does not guarantee, however, that \mathcal{P} will not become zero in the limit of $N \rightarrow \infty$. In fact, the requirement that

$$(5) \quad \lim_{N \rightarrow \infty} \frac{n_1^{(N)}}{N} > 0$$

is too strong even for productive affixes. On the other hand, productive affixes, but not unproductive ones, satisfy the requirement that

$$(6) \quad \lim_{N \rightarrow \infty} \frac{n_1^{(N)}}{V^{(N)}} > 0$$

(see Baayen 1989 for further discussion). For productive affixes, the number of hapaxes constitutes a non negligible portion of the total number of types, even for very large values of N .

Since \mathcal{P} is a function of N , and given the fact that \mathcal{P} becomes zero in the limit of $N \rightarrow \infty$ for both productive and unproductive affixes, we are forced to conclude that \mathcal{P} does not hand us the means for obtaining a measure of productivity that has a fixed value irrespective of sample size. However, such a measure can be obtained when we return to the original growth curve of V , and calculate an estimate of the number of types in the population.

2.2. The potential vocabulary size S

For productive word formation processes, the number of types in the population S , where S is defined as

$$(7) \quad \lim_{N \rightarrow \infty} V^{(N)},$$

is expected to be infinite, or at least larger than V by some significant factor. In the case of unproductive affixes, a finite value of S is expected that does not exceed V by much. Recalling that the shape of the frequency distribution of action nouns with vocalic alternation in Dutch already reveals that S must be finite and in fact quite small, we may proceed to ask whether it is possible to obtain an estimate of S on the basis of the grouped frequency distribution. The answer is yes, but to do so we have to make use of an additional assumption, namely, that some version of Zipf's law is valid for the underlying population.

As mentioned above, it is possible to write $V^{(N)}$ as a function of $V^{(M)}$ and $n_r^{(M)}$ given a sample of size M for which $V^{(M)}$ and $n_r^{(M)}$ ($r = 1, 2, \dots$) are known. This function is obtained on the assumption that each type is binomially distributed and occurs independently in the item sample.⁸ However, for technical reasons, this function does not lend itself very well to calculating $V^{(N)}$ for values of N which are very much larger than M . Although maximum likelihood techniques provide some insight into the upper and lower bounds of $V^{(N)}$ (Efron and Thisted 1976), more insightful results are obtained when we make the additional assumption that the n_r obey some version of Zipf's law.

Recall that the types which occur in some item sample can be ranked according to decreasing token frequency, as shown for simplex nouns in Dutch above. If f_i is the frequency of the i th type, then Zipf's law (Zipf 1935) states that

$$(8) \quad f_i \cdot i = K,$$

that is, the product of rank i and frequency f_i is a constant, for all i . Reformulated in terms of the n_r of the grouped frequency distribution, Zipf's law states that

$$(9) \quad n_r = C/r(r+1),$$

where the constant C is often equated with V , the number of types. In other words, Zipf's law specifies that the number of types occurring r times is a simple function of V and r . Unfortunately, this version of Zipf's law does not have general validity.

The problem with Zipf's law in the form given here is twofold. In the first place, it has been shown (Orlov 1983a, 1983b, Orlov and Chitašvili 1982a, 1982b, 1983a, 1983b, Chitašvili and Khmaladze 1989) that for a given sample Zipf's law (9) is valid only for some particular sample size, the so-called Zipf's size Z . In other words, it is not guaranteed that Zipf's law is accurate for a given item sample of arbitrary size. Often one will have to manipulate the size of the item sample N in order to obtain a reasonable fit to (9). For instance, (9) does not yield an accurate description at all of the empirical grouped frequency distribution of the 1927 tokens of action nouns with vocalic alternation in the EC. According to (9), n_r is a monotonically decreasing function of r , but Figure 2 shows that this is not the case for the sample size $N = 1927$, where n_r first increases and only then decreases. Figure 2 also shows that it is possible to obtain a somewhat better fit when the sample size is halved. Orlov and Chitašvili, who are the first to call attention to this remarkable state of affairs, take this factor of the sample size into account by enriching the model with an extra parameter $t = N/Z$, the factor by which the sample size N deviates from the Zipf size Z .

In the second place, Zipf's law has been found to be too simplistic. When plotted on double logarithmic graph paper, the graph of i and f_i should show up as a straight line. However, many samples show deviations, notably at the left hand and right hand ends of the curve. Various modifications and extensions of Zipf's law have been proposed, of which those by Mandelbrot (1962) and Simon (1955, 1960) are best known. The Waring—Herdan—Muller model (Herdan 1960, 1964, Muller 1979a, 1979b) is yet another example of a generalization of Zipf's law. Orlov and Chitašvili (1982a, 1982b, 1983a, 1983b) have shown that all these 'laws' are particular realizations of one general 'law' with three parameters. This generalized Zipf's law is, like Zipf's law itself, valid for only one particular sample size Z . When extended with the additional parameter $t = N/Z$, we obtain the so-called extended generalized Zipf's law. According to this law, the number of types for arbitrary sample size N , $V^{(N)}$, is proportional to the product of $V^{(Z)}$, the number of types for the Zipf size Z , and a function $F(\alpha, \beta, \gamma, t)$ that cannot be solved analytically for arbitrary values of the parameters α , β , γ and t . For $\alpha = \beta = \gamma = 1$, we have the extended version of the original Zipf's law (9). In this special case it can be shown that the potential vocabulary is infinite, that is, $\lim_{N \rightarrow \infty} V^{(N)} = \infty$. Hence the extended Zipf's law is a possible, perhaps a reasonable model for productive classes only. Unfortunately, the extended Zipf's law fails as a model for the frequency

distributions of the morphological categories I have studied: the theoretical values predicted for the n_r fail to provide anything near a satisfying match with the empirical values even in the case of the productive affixes (see Baayen 1989 for detailed discussion).

A generalization of Zipf's law which we have found to be particularly useful for the analysis of productive affixes is the so-called extended Yule—Simon law, which is obtained when the parameters α and γ of the function F of the generalized extended Zipf's law are set to unity. In this case $V^{(N)}$ can be expressed as a function of $V^{(Z)}$ and β that can be solved analytically, and considering the value of $V^{(N)}$ for $\lim_{N \rightarrow \infty}$ we find that

$$(10) \quad S = \begin{cases} V^{(Z)} \frac{\beta}{\beta-1} & \text{if } \beta > 1 \\ \infty & \text{if } 0 < \beta < 1 \end{cases}$$

In other words, for positive β smaller than 1, and for values of β not much larger than 1, S is infinite.

Table 4 lists the values of S obtained for the Dutch suffixes *-heid*, *-sel*, the suffix *-er*, which is used to form agent nouns (*gever* 'giver'), and the diminutive suffix *-tje* (*kamer-tje* 'small room'). For all grouped frequency distributions involved, a good fit is obtained: the theoretical values predicted for the model for the n_r , $r = 1, 2, 3, 4$, are very close to the empirical values.

Although the four suffixes of Table 4 are all productive, it is only *-tje* and *-er* which are characterized by infinite values of S , given (10). In the case of *-sel* and *-heid*, finite values are calculated. These values, however, exceed the number of types in the item samples by a substantial factor, given by

$$(11) \quad \mathcal{F} = S/V.$$

This factor \mathcal{F} is the inverse of Aronoff's (1976) index of productivity, the ratio of actual to possible words. The fact that the population number of types calculated for *-heid* and *-sel* is finite, even though these suffixes are productive, is probably due to the interference of non-linguistic factors on the global productivity of these affixes. Especially in the case of *-sel*, which forms patient nouns, the number of verbs which lend themselves for affixation with *-sel* is conceptually highly restricted. In other words, the value of S obtained from empirical samples for productive processes need not be infinite, since the characterization of productive processes in terms of infinite

Table 4. S , calculated on the basis of the extended Yule—Simon model

affix	t	β	V	S	$\mathcal{F} = S/V$
<i>-sel</i>	0.15	3.26	44	126	2.8
<i>-heid</i>	0.10	2.48	466	2063	4.4
<i>-er</i>	2.60	0.97	299	∞	∞
<i>-tje</i>	5.50	0.40	1031	∞	∞

S is based on considerations pertaining to the language system in the strict sense, without taking into account the possible influence of non-linguistic factors. However, whenever $S \neq 0$, the difference between V and S should be substantial for productive classes.

The extended Yule–Simon model appears to model the more productive affixes only. For less productive and unproductive classes, the parameter α of the general model should be allowed to take on values other than unity. The resulting model, for which only the third parameter, γ , is fixed at 1 is the extended version of the Waring–Herdan–Muller law.⁹ Using this model, we obtain the values of \mathcal{I} in Table 5 for some of the other Dutch morphological classes mentioned thus far (see Baayen 1989 for further details). Note that, as expected, \mathcal{I} assumes smaller values for these unproductive classes than for the productive affixes in Table 4. Hence, the index \mathcal{I} provides another way of measuring the productivity of word formation rules. Like \mathcal{P} , it correctly ranks morphological classes in an order of increasing productivity.

Turning to the interpretation of \mathcal{I} , it should be noted that we are dealing with a far more abstract notion than in the case of \mathcal{P} , our measure for the degree of productivity. The index \mathcal{I} is a ratio, not a probability. Moreover, it cannot be calculated directly from the empirical grouped frequency distribution, as in the case of \mathcal{P} . Instead, it is arrived at on the assumption that some version of Zipf's law is valid. Since we find it unlikely that intuitions concerning productivity arise from knowledge of the ratio of S on V , we are led to believe that \mathcal{I} should not be interpreted as a second measure of the degree of productivity of word formation rules. To our mind, \mathcal{I} is a measure of the potentiality of word formation rules, since it expresses the extent to which the number of 'actual' words in the corpus V exhaust the number of 'possible' words S . In the light of the fact that S may assume finite values for productive processes, even though the calculus of word formation predicts that S should be infinite here, we will refer to \mathcal{I} as a measure of 'pragmatic' potentiality to emphasize the fact that various pragmatic and conceptual factors codetermine the value of S and hence \mathcal{I} .

2.3. The actual vocabulary size V

We are left with the linguistic interpretation of the vocabulary size V , the number of types in the item sample. Since \mathcal{P} is the growth rate of V for sample size N , and since S is obtained by considering the growth curve of V

Table 5. S , calculated on the basis of the Waring–Herdan–Muller model, with t fixed at 1

	V	S	\mathcal{I}
action nouns with vocalic alternation	100	109	1.09
simplex nouns	1495	1955	1.31
-te	39	58	1.49

in the limit for $N \rightarrow \infty$, the interpretation of V itself requires further thought. Moreover, the fact that word formation rules are represented by different numbers of types within the same frame sample is in some sense relevant to the general issue of productivity. Consider Table 6, which lists V and \mathcal{P} for a number of affixes in Dutch. Even though all affixes of Table 6 are productive, they show up with widely varying numbers of types in the EC. At the one extreme we have the suffix *-erd*, which is represented by only six types (and nine tokens).¹⁰ At the other extreme we find the nominal compounds, for which the probability of coming across new types is roughly one on five, at a point where some 4000 types have already been sampled.

The differences then in the numbers of types seem to reflect the extent of use of these morphological processes.¹¹ As was mentioned in Section 1, the suffix *-erd* has a low extent of use, especially in written language, a fact which may explain the low number of types in *-erd*. Nominal compounds, on the other hand, enjoy a tremendous extent of use that is at least in part due to their semantic versatility (Downing 1977) and the wide range of styles in which they can be put to use. Table 6 also illustrates that affixes with highly similar values of \mathcal{P} , such as *-sel* and *-er* ($\mathcal{P} = 0.08$) may differ as to their extent of use. Patient nouns in *-sel* are represented by 44 types, the agent nouns in *-er* by 299 types. In a global sense then, the latter affix is more productive than the former.

The 'global productivity' of a number of English word formation processes is summarized in Figure 3, with the degree of productivity \mathcal{P} on the horizontal axis and the extent of use V on the vertical axis.¹² Typically unproductive affixes, such as the prefix *en-* (*enchain*, $\mathcal{P} = 0$) are found in the lower left hand corner of the plot, while typically productive affixes such as *-ness* ($\mathcal{P} = 0.0044$, $V = 497$) are characterized by large numbers of types in combination with high values of \mathcal{P} .

In general, it is unclear how to evaluate the different contributions of \mathcal{P} and V to the global productivity when arbitrary affixes are compared. It is only for a number of special cases that it is fairly evident how to proceed. In the case of rival affixes, such as *-ness* and *-ity*, we may assume that the non-linguistic factors underlying the extent of use are roughly identical (but see Riddle 1984). Here we cannot trace differences in the number of types to

Table 6. V and \mathcal{P} for selected productive affixes of Dutch

	V	\mathcal{P}
<i>-erd</i>	6	0.444
<i>-ster</i>	30	0.231
<i>-sel</i>	44	0.080
<i>-er</i>	299	0.076
<i>-heid</i>	466	0.114
<i>-tje</i>	1031	0.253
N + N compounds	4277	0.225

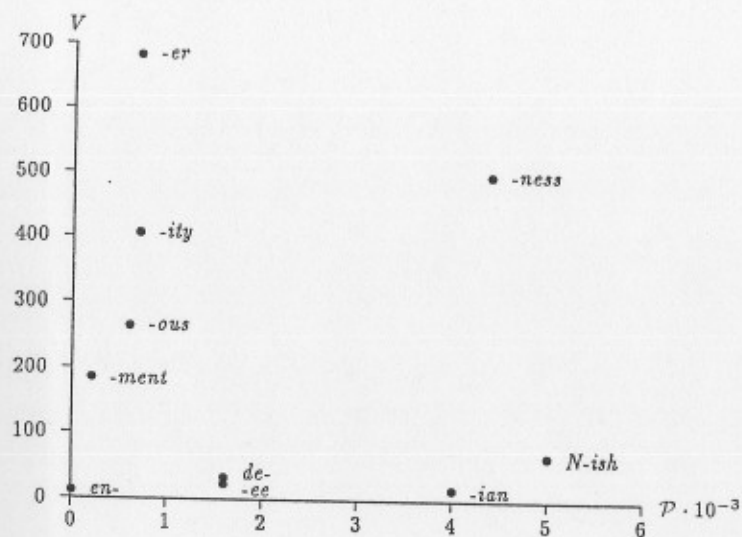


Figure 3. Global productivity of a number of English word-formation processes as found in the Cobuild corpus. The degree of productivity \mathcal{P} is plotted on the abscissa, the extent of use V on the ordinate.

differences in the extent of use. Instead, such differences reflect the extent to which the base words which satisfy the conditions on the relevant rules have been used, while differences in \mathcal{P} relate to differences in the extent to which the remaining available base words can be used to create neologisms. In the case of affixes with roughly similar \mathcal{P} , for instance, *-ness* and *-ian* in Figure 3 ($\mathcal{P} \approx 0.004$), the difference in global productivity is conditioned by a sizeable difference in the extent of use (491 versus 16 types respectively). When affixes have roughly identical extent of use, as is the case for *-ian* and *-ee* (27 and 23 types respectively), the difference in global productivity arises from a difference in the degree of productivity. Unfortunately, when both extent of use and degree of productivity are different for non-rival affixes, a ranking in terms of global productivity cannot be obtained on the basis of \mathcal{P} and V . However, we can study what happens to the positions of the affixes in Figure 3 when we send N to infinity. As was discussed above, this has the effect that \mathcal{P} becomes zero for all affixes, irrespective of their productivity. Hence a plot is obtained in which all affixes are positioned on the ordinate. Interestingly, their positions on the ordinate represent their respective values of S , since S is the value of $V^{(N)}$ for $N \rightarrow \infty$. Consequently, the global productivity of word formation processes can be measured indirectly in terms of their pragmatic potentiality \mathcal{S} .

Summing up, by focusing on the growth curve of V , we have obtained two complementary techniques for evaluating the global productivity of word formation processes. By jointly considering the degree of productivity \mathcal{P} and the extent of use V , the global productivity of a word formation process

can be evaluated on the basis of the number of tokens, the number of types and the number of hapaxes in the item sample. This technique has the advantages of computational simplicity and generality. It has the disadvantage that it does not provide an explicit ranking of affixes according to their global productivity on a simple (monodimensional) scale. The second technique allows us to obtain a single explicit ranking by means of the index of pragmatic potentiality, \mathcal{S} . This technique has the disadvantage that the computation of S and hence \mathcal{S} involves some numerical calculations which cannot be carried out by hand. Moreover, it assumes the validity of some version of Zipf's law, an assumption that need not be met.¹³

3. FREQUENCY, PRODUCTIVITY AND THE MENTAL LEXICON

In the preceding section we have argued that productivity and token frequency are correlated. This finding is of interest for the theory of the mental lexicon, since frequency information is recorded in memory. Hasher and Zacks (1984: 1379) point out that in general

The processing of frequency of occurrence information is remarkable. Information about frequency is recorded in memory without a person's intention to do so. The information stored in this way is apparently no less fine-grained than is the information stored when intention is operating. Training and feedback do not improve the ability to encode frequency information. Unlike virtually every other cognitive skill examined in the history of the field, memory for frequency shows a developmental invariance from early childhood through young adulthood to middle and old age. Similarly, there are no effects of differences among people in motivation, intelligence, and educational background. The processing of frequency information is unaffected by reductions in cognitive capacity stemming from depression, old age, or multiple task demands.

Word frequency represents a particular instance of frequency information that is unintentionally accumulated in memory, and its effect in various experimental tasks is well known. Word frequency affects the signal to noise ratio under which stimuli can be understood. For instance, Rubenstein and Pollack (1963) report that when words are presented in noise a reduction in the signal to noise ratio of 3–4 dB should be balanced by a tenfold increase in word frequency in order to maintain a given level of intelligibility. Word frequency also affects response latencies in the lexical decision task (Whaley 1978), the naming task (Forster and Chambers 1973) and the classification task (Monsell 1985), and it has also been found to be correlated with parameters of eye movements, such as fixation durations (Rayner and Duffy 1986). Models of lexical access account for this pervasive effect of word frequency in various ways. For instance, the serial search model (Forster 1976, Taft 1988, Bradley and Forster 1987) accounts for the frequency effect by modelling lexical access on a serial search through a frequency-ordered list. Other models encode the frequency effect in counters associated with the lexical representations of the types in memory. For instance, Marslen-Wilson (1987) codes word frequency into the activation level of the lexical representations, allowing the activation of high-frequency word candidates in the cohort to rise more rapidly than the activation of low-frequency

candidates. Recent work on 'subsymbolic' models of lexical access locates the frequency effect in the connection weights of the distributed representations (Rumelhart and McClelland 1986, McRae, Jared and Seidenberg 1990). We will take a conservative stand and assume that the token frequencies of the types encountered in experience affect the activation level or the representational strength of the lexical entries of these types in memory.

Under this assumption, the differences in the frequency distributions observed for productive and unproductive word formation rules correspond with differences in the distributions of the activation levels of the lexical representations. Unproductive word formation rules are characterized by lexical representations with generally high activation levels, while large numbers of lexical representations with very low activation levels are typical of productive word formation rules. We may note that the high token frequencies of unproductive formations are functional in the light of the fact that the linguist's 'unproductive rule' is not an active rule in the mental lexicon. Unproductive rules summarize patterns of regularities. They have no generative power, as shown by the fact that unintentional, spontaneous neologisms cannot be formed by means of such rules (Schultink 1962, Uhlenbeck 1977). Consequently, the high token frequencies of unproductive formations guarantee that they can be efficiently recalled from memory. In contrast, productive formations are backed up by the presence of rules with psychological validity. Such formations will generally be stored in memory, but, especially in the case of types with very low frequencies of occurrence, storage in memory is not obligatory as long as all properties of the types are fully predictable by rule. When not available from memory, for instance because of decay over time of the lexical representation, such low-frequency formations can be parsed or generated by the relevant word formation rules of the language.

Summing up, the relative dominance of high-frequency types in the frequency distributions of unproductive word formation rules can be understood to guarantee the efficient retrieval from memory of formations for which, as is the case for simplex items, no word formation rules are available. The large numbers of rare types in the frequency distributions of productive word formation rules suggest that productive rules operate in parallel with a memory-based access procedure, securing efficient access for those formations for which the representational strength is insufficient for the memory-based access procedure to complete successful retrieval.

We have argued that productive word formation rules guarantee that low-frequency formations such as the hapaxes can be processed in case the memory-based access procedure fails to do so. This possibility is given with the fact that such rules have to be available anyway for the processing of new complex formations that have not been encountered previously. We may press our argument one step further, and consider the possibility that productive word formation rules speed up the processing of low frequency formations. In that case, two strategies are available for retrieving existing words from memory, a relatively slow rule-based access procedure, and a relatively fast memory-based access procedure, which operate in parallel.

High-frequency types, irrespective of whether the corresponding word formation rule is productive or not, are efficiently accessed by the memory-based address procedure. For such types, no benefit from the rule-based address procedure is to be expected, since access by memory will have been completed before access by rule. In the case of low-frequency items, the speed of retrieval might well benefit from statistical facilitation. For such items, the memory-based access procedure operates more slowly than for high-frequency words. Hence, it is for these items that an effect of word formation rules on the speed of processing is most likely to be felt.

The analysis given here makes essential use of a number of assumptions, to which we now turn. In the first place, for our analysis to be valid, the frequency effect should be due only to the frequency of words in experience. This amounts to accepting the validity of the so-called principle of acoustical equivalence (Morton 1968: 22, see also Broadbent 1967 and Broadbent and Broadbent 1975), according to which

No information as to the frequency interval of the word can be gained from the stimulus and any word is likely to be confused on the basis of its stimulus properties with a word of any of the frequency intervals.

If the principle of acoustical equivalence is valid, the effect of word frequency on response latency is maximal: the stimulus itself provides no clue as to its frequency range. If, however, information concerning its frequency interval can be obtained from the stimulus it may well be that this information, either wholly or in part, is the determinant of response times rather than word frequency itself. Landauer and Streeter (1973) argue for the latter position. One of their arguments against the principle of acoustical equivalence concerns the phonemic and graphemic characteristics of high versus low frequency words. For instance, the phonemes *n*, *l* and *t* are characteristic of the more common words while *z*, *p* and *g* are favoured by the rare words of their study, so that there are reliable differences in the distribution of phonemes between the two frequency sets. Similarly, Pisoni *et al.* (1985: 85) have found that high frequency words tend to be composed of consonants having an alveolar place of articulation and seem to disfavour those consonants having a velar place of articulation. They suggest (1985: 85) that "frequently used words may have succumbed to pressures over the history of the language to exploit consonants that are in some sense easier to articulate".

Landauer and Streeter (1973: 120) also points out that the more common words have larger similarity neighbourhoods, a similarity neighbourhood being defined as

the set of words in the language from which a given stimulus word is indistinguishable after a specified loss of information about the stimulus word.

Landauer and Streeter (1973) operationalized this definition by equating a letter with 'a specified loss of information', and considered the neighbours, that is, the words that shared the same letters in all but one position, of a set of common and rare-four-letter words in the Kučera and Francis (1967)

word list. They found that (1) common words have larger similarity neighbourhoods than rare words, and (2) that the words in the similarity neighbourhoods of high-frequency words tend to be high-frequency words too. At first sight, this result suggests that common words should be less easily perceived than rare words since frequent words have more neighbours with which they can be potentially confused. This line of reasoning leads to the paradox that high-frequency words should take more time in recognition than low-frequency words, contrary to fact.¹⁴ One possible explanation of this paradox can be formulated for interactive activation models. In such models, highly similar words can profit from their large degree of similarity and jointly acquire higher activation levels before the final differentiation stage, the so-called 'gang-effect' (Nusbaum 1985: 455). Alternatively, high-frequency words may contain the perceptually more salient phonemes. Perceptual salience would then undo the negative effect on processing of dense neighbourhood structure (Landauer and Streeter 1973).

However, the frequency of the neighbours relative to the stimulus word has been argued to be more important than the size of the similarity neighbourhood. For instance, Grainger *et al.* (1989) claim on the basis of a visual lexical decision task and on the basis of measurements of gaze durations that when factors such as word frequency, experiential familiarity and orthographic characteristics (bigram frequencies) are strictly controlled for, the fact that a target word has orthographic neighbors of higher frequency than itself has the effect of increasing the duration of the lexical processing of this target word. Luce (1986: 24) develops a neighbourhood probability rule which expresses the probability of retrieving the correct stimulus representation from among its neighbours on the basis of the probability of identifying the stimulus in its neighbourhood, the probabilities of confusing neighbours with the stimulus, and the frequency weights of both stimulus and neighbours. By means of this rule he is able to account for the fact that, although on average high-frequency words are identified more accurately than low-frequency words, high-frequency words residing in dense, high-frequency neighbourhoods are identified less accurately than low-frequency words residing in sparse, low-frequency neighbourhoods.

The arguments put forward by Landauer and Streeter (1973) and subsequent work on lexical density present a strong case against word frequency as a fully independent factor in lexical access. On the other hand, Gardner *et al.* (1987) show that at least some portion of the word frequency effect is due simply to frequency of occurrence in experience. They required subjects of two occupational groups, nurses and engineers, to make lexical decisions about two sets of occupationally related words, controlling for overall frequency by means of the Kućera-Francis (1967) word count. They found that the response times of the nurses to medical words were shorter, while the engineers responded more quickly to engineering words. Thus the same words with the same segmental make-up lead to significantly diverging response times as a function of occupational background. They argue (1987: 28) that this difference in occupational background and work experience

is the heart of a true frequency based difference. Word frequency, in an approximate way, reflects the familiarity of the subject with the meaning of a word and the contexts in which it is likely to occur, and may also indicate how recently it has been seen.

They conclude that word frequency is a separate factor in lexical access, with the proviso that the magnitude of the effect is in all likelihood not a function of word frequency alone.

These results show that the assumption that empirical frequency distributions are isomorphic with distributions of activation levels is too simplistic. First, high frequency formations may enjoy processing advantages over low frequency words on the basis of their segmental constituency. Second, the presence of a dense high-frequency neighbourhood structure may slow down lexical processing. Such processing advantages and disadvantages cannot be accounted for in terms of activation levels. To this we should add the fact that the frequency of the stem of a derived word has been found to co-determine response latencies in the lexical decision task (Taft 1979, Laudanna and Burani 1985, Burani and Caramazza 1986, Cole *et al.* 1989). For stimuli with identical whole word frequencies, shorter response latencies have been observed for the formations coined from base words with the higher token frequencies. At the same time, response latencies to words coined from bases with identical frequency have been found to decrease with increasing whole word frequency.¹⁵ In other words, formations with high-frequency base words enjoy processing advantages over formations of similar token frequency with low-frequency base words. Again, this difference in processing advantage cannot be expressed in terms of activation levels.

A second assumption underlying the analysis presented at the beginning of this section concerns the locus of the frequency effect. Balota and Chumbley (1984, 1985) report a series of experiments which seem to indicate that task-specific stages following lexical identification are highly frequency-sensitive rather than the stage of lexical identification itself. If correct, their results seriously question the idea that word frequency is coded into the activation level of lexical entries. They suggest that the role of word frequency in the mental lexicon is primarily linked with postaccess processes (for instance, pronunciation assembly and response execution in the naming task), rather than with the representational strength of words in memory. This would imply that word frequency is of relatively low importance with respect to how well words are represented in memory. Consequently, the presence or absence of complex words in memory would be relatively independent of frequency, seriously questioning the analysis proposed here. However, the claim that word frequency has minimal effects on lexical identification has in turn come under attack. Results obtained by Monsell *et al.* (1989) strongly argue against the view that lexical lookup is not intrinsically frequency-sensitive.

A third assumption underlying the present analysis is that both productive and unproductive formations are stored in memory. In other words, we assume that for languages as Dutch and English words are stored in and

retrieved from the lexicon as full forms, with the proviso that rules may be involved in making available the representation of very low frequency (productive) complex forms. This assumption comes close to what is known as the full listing hypothesis (Butterworth 1983), according to which all words encountered in experience are stored in memory. Other theories argue that morphological rules are necessarily involved in lexical access. For instance, in the serial search model incoming stimuli are stripped of their prefixes before they are subjected to a sequential comparison with the initial syllables in the access files. The augmented addressed morphology model (Laudanna and Burani 1985, Caramazza *et al.* 1988) claims that both full forms and affixes are available in memory, that memory-based retrieval is relatively fast and that rule-based retrieval is relatively slow.

Unfortunately, the evidence for the operation of morphological rules in lexical access is far from unequivocal. For instance, experimental results in favour of prefix stripping (Taft and Forster 1975, Taft 1979, 1988, Lima 1987) are balanced by studies which report experimental results that are in conflict with this model (Manelis and Tharp 1977, Tyler *et al.* 1988, Cole *et al.* 1989). Furthermore, a single model need not be valid for different languages. Results obtained by Jarvella *et al.* (1987) and Schreuder *et al.* (1990) suggest that the augmented addressed morphology model may well be correct for Italian, but not for Dutch. The former language has a richer morphology than the latter. As pointed out by Schreuder *et al.* (1990: 14),

This means that the amount of storage saved by only storing stems and affixes of verbs and not storing all full forms would be much larger for Italian than for Dutch. Thus, if a difference emerges when these languages are investigated in psycholinguistic experiments, then one would expect effects indicating a decomposed lexicon for Italian verbs but not for Dutch verbs.

The only evidence for word structure in Dutch obtained by Schreuder *et al.* (1990), Schreuder (1991) concerns verbs with separable particles, which have a phrasal rather than a word status. These formations can occur distributed over the clause, hence the the syntax enforces some form of decomposed storage here.

Restricting the discussion to Dutch and English, languages for which the full listing hypothesis appears to be a reasonable model, the question arises whether it can be maintained that rules are involved in retrieving the lexical representations of productive complex words with very low frequencies. To our mind, experimental evidence does not argue against this hypothesis. Experiments designed to obtain evidence for the use of rules in the lexical access of existing words typically make use of stimuli of moderate length and not too low frequency. Given the analysis proposed here, the implication is that the experimental design forces research to focus on those formations for which evidence for rules is unlikely to be obtained. Consider Figure 4, which plots extent of use and degree of productivity for Dutch simplex verbs, for simplex verbs prefixed with *be-* (*be-la(a)d-en* 'to load with'), for adjectives in *-baar* '-able' coined from these verbs with *be-*, and finally for abstract nouns in *-heid* coined from these adjectives. Figure 4 illustrates that an increase in

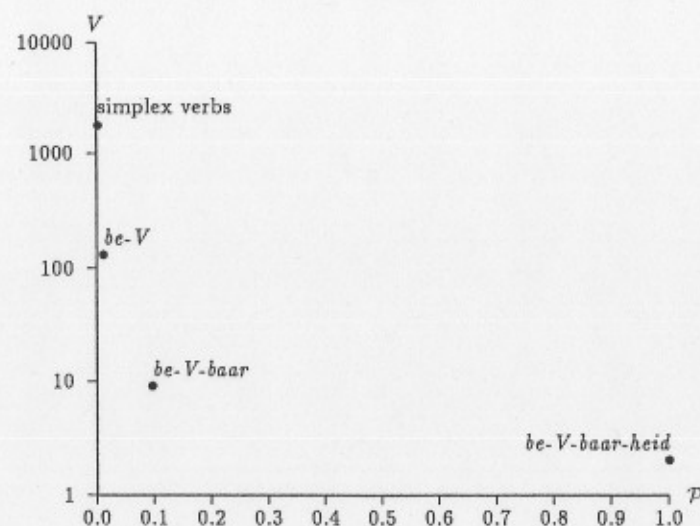


Figure 4. Affix stacking in the $\mathcal{P}-V$ plane.

morphological complexity corresponds with a fall in the number of types, and with a rise in the value of \mathcal{P} . The highest degrees of productivity are measured for those formations which make the most extensive use of the morphological possibilities of the language.¹⁶ Consequently, if rules are involved in lexical access, their effect should be measurable for multiply complex formations. In contrast, complex words derived from simplex bases will generally have high token frequencies. Here the effect of morphological rules will be minimal. Unfortunately, it is precisely these words which figure prominently in psycholinguistic experiments designed to trace the effect of word formation rules in perception. Hence the possibility that word formation rules may aid the retrieval of very low-frequency regular complex words is not ruled out by experimental results.

This conclusion is in line with results obtained by Stemberger and MacWhinney (1986, 1988) for the production of inflectional endings. They argue on the basis of studies of speech errors in natural and experimental situations that words with irregular flexional endings and high-frequency regularly inflected words are stored. Given that a high token frequency safeguards formations against certain kinds of production errors, the fact that low-frequency regular verbs show up with more speech errors than high-frequency regular verbs can be explained.

A final assumption underlying our initial analysis should be made explicit here, namely, the idea that memory-based address has a real-time advantage with respect to the speed of processing above rule-based address (Laudanna and Burani 1985, Caramazza *et al.* 1988, Meys 1985, MacWhinney 1978). Evidence in support of this claim concerns the observation that neologisms require longer response latencies than well-established, existing words in for instance the lexical decision task.

The examination of the assumptions underlying our initial analysis has shown that frequency distributions and distributions of activation levels are not isomorphic. Although it can be maintained that unproductive formations are especially well-suited to being processed by the memory-based access procedure in the light of their high token frequencies, an analysis of the relative processing advantages with respect to memory-based and rule-based access of productive and unproductive formations in terms of words frequency only is not complete. In the first place, recall that the cumulative root frequency co-determines the speed of lexical access. This is of interest to the issue at hand, since unproductive affixes are typically found attached to higher frequency base words than productive affixes, as pointed out by Aronoff (1982) for English *-ness* and *-ity* and Baayen (1989) for Dutch *-heid* and *-te*, as shown in Table 7.¹⁷ Consequently, the frequency distributions of the derived words underestimate the extent to which unproductive formations enjoy a processing advantage over productive formations in terms of the memory-based access route.

In the second place, productive formations are the most vulnerable to decay of representational strength over time, since they have the lower token frequencies. Hence productive formations may depend on the availability of word formation rules to a higher extent than suggested by their frequency distributions.

In the third place, it should be noted that while unproductive formations are characterized by processing advantages with respect to the memory-based address procedure, they may have a disadvantage with respect to the rule-based address procedure. In perception, a parsing disadvantage arises when the number of tokens with a 'pseudo-affix' is large with respect to the number of tokens with the true affix. As a case in point, consider the number of tokens that end in the suffix *-heid* in the Dutch INL corpus (40 000 000 word forms) and the number of tokens that do not end in *-heid*, but which in isolation are phonologically indistinguishable from the true suffix, as listed in Table 8. We find that confusion will arise only sporadically. Interestingly, the reverse holds for its unproductive rival *-te*. When a string in 'heid' is presented, one may be 99% sure that it ends in the suffix *-heid*. For strings in 'te', the suffix for abstract nouns is involved in only 3% of the occurrences (compared e.g. *zwakte* 'weakness' with *zakte* '(he) failed').

Summarizing our discussion, we arrive at the hypothesis that unproductive formations occupy those niches in lexical space where they are maximally sustained by the mechanisms which underly memory-based access. Con-

Table 7. Adjectives underlying *-te* and *-heid* (EC)

	<i>N</i>	<i>V</i>	n_0	n_1	n_1/N
<i>-te</i>	5663	43	2	2	0.0004
<i>-heid</i>	16191	466	74	49	0.0030

Table 8. Possibility of confusion with homonym strings for Dutch *-te* and *-heid*. Data from the Celex INL database of 40 000 000 word forms

affix	tokens with affix	tokens without affix	total	fraction correct
<i>-te</i>	54539	1593903	1648441	0.03
<i>-heid</i>	194349	1637	195986	0.99

versely, productive formations are not particularly geared to being processed by the memory-based access procedure.

We conclude this section with a final remark. We have discussed Schreuder *et al.*'s (1990) observation that the augmented addressed morphology model may be right for Italian, but that it makes the wrong predictions for Dutch. Their explanation for this state of affairs is that for Italian memory capacity would be strained in the case of full listing, given the rich morphology of this language when compared to Dutch. To this we may add that the richness of the Italian inflectional morphology has the effect that where Dutch uses a single inflectional form, Italian makes use of several distinct forms. Consequently, the token frequencies of each of these Italian forms will be lower than the token frequency of the single Dutch form. In turn, lower token frequencies raise the degree of productivity of the relevant inflectional rules and, if the analysis presented here is correct, the probability that rules are involved in lexical access, a conclusion that is in harmony for the results obtained with the augmented addressed morphology paradigm for this language.

4. TOKEN FREQUENCY, LEXICAL REPRESENTATION AND MORPHOLOGICAL RULES

In the previous section we have considered the processing advantages and disadvantages of productive and unproductive formations by focussing on the way existing formations are stored in the mental lexicon. The aim of the present section is to consider the relation between token frequencies and morphological rules. Three morphological models which recognize the importance of token frequencies are reviewed, namely Anshen and Aronoff (1988), Bybee (1985, 1988) and Rumelhart and McClelland (1986). The question with which we will confront these models centers on the way in which word formation rules express the fact that their degree of productivity may vary with the kind of input formations they apply to. For example, when we consider the English suffixes *-ness* and *-ity* as they attach to base words in *-ive* and *-ible*, we find (see Table 9 and Aronoff (1976, 1982) and Anshen and Aronoff (1988) for further discussion) that *-ness* is productive after *-ive* but wholly unproductive after *-ible*, while *-ity* is productive after *-ible* but far less productive after *-ive*. As pointed out in Aronoff (1982), the formal properties of these rules do not give any indication of the way in which the degree of productivity of say *-ity* varies with the kind of the base word it

Table 9. Productivity statistics for *-ness* and *-ity* for selected input types, as found in the Cobuild corpus.

suffix	V	P
<i>Xivity</i>	18	0.0003
<i>Xiveness</i>	27	0.0086
<i>Xibility</i>	36	0.0008
<i>Xibleness</i>	0	0.0000

attaches to. How then should these differences in productivity be accounted for? Anshen and Aronoff (1988) do not explicitly address this issue, but they develop an interesting theory of the mental lexicon in which both 'rules that look much like those written by linguists' and that are part of a speaker's competence (1988: 648) and token frequencies play a rôle. Their analysis is framed within a model of the mental lexicon that is rather similar to the one we have proposed in the previous section. In their model, formations are accessed by rote, by rule or by analogy, following MacWhinney (1978). Of these three ways of dealing with new words, analogy is argued to be slower than rote and rule, and rote is said to be faster than rule and analogy. The three methods of obtaining lexical items are assumed to operate in parallel.

Within this model Anshen and Aronoff (1988) try to account for differences in productivity in terms of blocking, along the lines of Aronoff (1976). First consider the irregular plurals of English, which block the actuation of the regular plural forms. The token frequencies of the irregular plurals in the Cobuild corpus are listed in Table 10. Given that the high-frequency irregular plurals are stored in the mental lexicon, and given the hypothesis that retrieval from memory is completed more quickly than the process of building a regular plural in *-s*, the coining of regular plurals is blocked by the existing high-frequency irregular plurals, which are generally accessed before the singular base word becomes available as input to the plural rule: six of the ten irregular plurals in Table 10 have a higher frequency than the corresponding singular form. Rainer (1988) develops a theory of blocking along similar lines on the basis of Italian and German data. He shows in some more detail that the force with which an item blocks roughly synonymous formations is a function of the frequency of the blocking word.

Anshen and Aronoff (1988) interpret the differences in mean token frequency observed for *-ness* and *-ity* in Aronoff (1982) in a similar vein. There are two sets of facts that they attempt to explain. First, high-frequency *glory* blocks **gloriosity*, as predicted by their theory of blocking. However, *glory* does not block *gloriousness*, and similarly *curiousness* appears to be well-formed, even though *curiosity* exists. Formations like *gloriousness* and *curiousness* run counter to what the blocking principle predicts. Anshen and Aronoff solve this problem by claiming that there is sufficient overlap in processing time of memory-based and rule-based access to allow the simul-

Table 10. Singular and plural: token frequencies in the CC (18 000 000)

	singular	plural
<i>child</i>	7619	11656
<i>die</i>	1537	223
<i>foot</i>	2079	4130
<i>goose</i>	114	85
<i>louse</i>	22	41
<i>man</i>	19136	11800
<i>mouse</i>	146	182
<i>ox</i>	79	36
<i>tooth</i>	242	1391
<i>woman</i>	6034	9038

taneous existence of rival forms (1988: 648). On the basis of this claim they argue that "it is reasonable to assume that a person needing an abstract noun semantically related to an *-ous* adjective may either use an already existing lexical item in *-ity* or *-y* or build a new *-ness* form by rule." (1988: 652)

The second set of facts requiring an explanation concerns the supposed blocking relation of *-ness* and *-ity*. Recall that the irregular plurals generally have higher token frequencies than their base words, causing them to block the actuation of the regular plural formations. In the case of *-ness* and *-ity* the token frequencies of the derived words are lower than those of the underlying base words, but in the case of *-ity* the base word frequencies are generally higher than in the case of *-ness*. This suggests that formations in *-ity* have two processing advantages with respect to formations in *-ness*: their high frequencies guarantee storage in the mental lexicon, and the relatively high frequencies of their base words guarantee rapid lexical access. With respect to *-ness*, Anshen and Aronoff (1988) argue that, rather than being stored with their bases as in the case of *-ity*, such formations are not stored at all but constructed by rule as needed. Their argument builds on the observation that the mean token frequency of words in *-ness* is less than that of words in *-ity*, both in a production test and in corpus-based frequency counts. Anshen and Aronoff are now in a position to argue that the stored formations in *-ivity* block the actuation of the rival formations in *-ness*, but they are left with the problem how to explain the fixed choice for the established formations in *-ness*, formations which they must assume are generated by rule for each instance of use and never stored. However, according to Anshen and Aronoff, this fixed choice is inevitable since "for *x-iveness* words the relatively long time necessary to get to a (supposed) lexical entry in *-ivity* means that the rule-based *-iveness* form is created first, blocking access to and thus the existence of a lexically based *-ivity* form" (1988: 653). In other words, without being stored in memory, *-iveness* words are claimed to block the rival formations in *-ivity*.

This analysis is the psycholinguistic version of the blocking theory devel-

oped in Aronoff (1976). As has been pointed out by for instance van Marle (1985) and Rainer (1988), Aronoff's (1976) blocking theory is inconsistent. It is highly circular to argue that 'possible' words in *-iveness* are blocked by 'actual' words in *-ivity* and at the same time claim that 'possible' (but not 'actual' words) in *-iveness* block the actuation of 'possible' words in *-ivity* where neologisms and established formations in *-ness* coined from base words in *-ive* are concerned. Rainer (1988) argues convincingly that the competence theory notion of listing does not carry over to theories of the mental lexicon. Nevertheless, this is what Anshen and Aronoff (1988) attempt to do. Not surprisingly, this attempt fails. In fact, the claims (i) that formations in *-ness* are not stored in memory and (ii) that there is a sizeable overlap in processing time for the memory-based and the rule-based access procedures are unconvincing. The first claim ignores the evidence discussed in the preceding section that at the least the higher-frequency regular formations are stored in the lexicon. It also ignores the simple fact that the more frequent types in *-ness* have token frequencies of a magnitude similar to those of formations in *-ity*, and the fact that derivation often involves concept formation (Meys 1985). For instance, it is unlikely that a high-frequency word like Dutch *snelheid*, 'quick-ness', which in English has the simplex counterpart *speed*, is re-invented for each instance of use. Except for the lowest frequency ranges, we must assume that formations in *-iveness* are stored in the mental lexicon, and that they block the actuation of the rival forms in *-ivity*.¹⁸ In this way we can avoid the inconsistency of claiming (1988: 653) that *-ivity* words are accessed rapidly when they have to block *-iveness*, and slowly when *-iveness* has to block *-ivity*.

Turning to the second claim, it should be noted that the introduction of an overlap in processing time for 0 frequency *curiousness* and *gloriousness* and 410 frequency *curiosity* and 419 frequency *glory* (Cobuild corpus) amounts to the claim that consistent differences in processing speed for memory-based and rule-based address arise only for words with rather large differences in token frequency. Apart from the fact that this claim, which pronounces differences in frequency irrelevant for a sizeable frequency range, is unlikely to survive experimental verification, we may note that it predicts that a substantial portion of the established formations in *-iveness* should have well-formed rival formations in *-ivity*, contrary to fact. The fact that *curiosity* does not block *curiousness* in principle, together with the fact that rival *-ness* — *-ity* pairs are extremely rare, should be explained not in terms of frequency, but in terms of the subtle differences in the semantics of these suffixes (Riddle 1984).

When we modify Anshen and Aronoff's (1988) theory by dropping these two incorrect claims, we obtain a model in which we have, on the one hand, lexical representations of varying representational strength, and on the other, formal word formation rules that apply to these lexical representations. The differences in productivity observed for *-ness* and *-ity* for base words in *-ive* and *-ible* can be accounted for by formulating the formal rule for *-ness* in such a way that base words in *-ible* are excluded from its derivational domain, and by similarly removing base words in *-ive* from the derivational

domain of the *-ity* rule. Although this version of the model is correct as far as the actual use of these abstract nouns in the Brown and Cobuild corpora is concerned, it does not shed light on the fact that Walker (1936) is able to generate 27 formations in *-ibleness* for his rhyming dictionary, and that speakers are able to come up with 9 of these formations plus 12 new ones in the production test of Anshen and Aronoff (1988), and that in the same production test subjects created 9 neologisms in *-ivity* (see Table 11). Since base words in *-ive* and *-ible* are excluded from the input domains of the rules of *-ity* and *-ness* respectively, these neologisms cannot be explained by this modified version of their model.

Let us, for the sake of the argument, assume that the experimental situation of Anshen and Aronoff's (1988) production task did not force subjects into relying on analogical formation rather than formation by rule. In that case, we may seek to adjust the model by assigning the *-ness* and *-ity* rules probabilities of application. The problem with this modification is, unfortunately, that the introduction of probabilistic word formation rules in a theory of lexical processing should preferably be avoided. There is a logical problem, a probabilistic rule can be applied only when it has been ascertained in some non-probabilistic way that the rule has to apply probabilistically, and there is an efficiency problem, since for instance in perception a neologism in *-ness* may fail to be parsed, necessitating repeated application of the probabilistic rule in order to obtain the required parsing. However, when rival word formation rules are available for the production of neologisms, a choice has to be made, and it is likely that this choice is guided by the degrees of productivity of the rival rules involved. Instead of assigning probabilities of application to word formation rules, we might therefore argue that word formation rules should be enriched with indices of their degree of productivity, which may serve to guide affix choice in the case of rival word formation processes.

In turn, this proposal is challenged by the problem that it is unclear how to choose this index of productivity. Given the data of Tables 11 and 9, we are confronted with the following problem. If we assign the relevant word formation rules indices of productivity that generalize over all input domains, for instance, $\mathcal{P} = 0.0044$ for *-ness* and $\mathcal{P} = 0.0007$ for *-ity*, we can explain why *-ibleness* shows up with more nonce formations in Table 11 than *-ibility*

Table 11. Type counts in Anshen and Aronoff's (1988: 645) production test

input domain	extant words	neologisms
<i>-ivity</i>	19	9
<i>-iveness</i>	17	16
<i>-ibility</i>	38	8
<i>-ibleness</i>	9	12

in the production task. On the other hand, frequencies of (normal) use indicate that *-ibleness* is not a natural option. In fact, *-ness* is completely absent in this derivational domain, while, as shown by the value of \mathcal{P} of *-ity* for base words in *-ible*, *-ity* is slightly stronger in productivity here than for the *Xive* input domain.¹⁹ That *-ity* has its focus of productivity for *Xible* base words, and has ousted *-ness* here, cannot be obtained from the overall productivity values of *-ness* and *-ity*. Evidently, both the overall productivity index and the indices for particular input domains are relevant, but how the one and the other interact, and how this interaction can be formally represented in the rule, remains unclear.

This state of affairs is typical of Item and Process models (Hockett 1954), in which the rules and the representations are strictly separated. The frequency aspect of productivity is intimately linked with the lexical representations, the way these 'Items' of the model are stored and the way in which they interact in the mental lexicon. However, while phonological, morphological, syntactic and semantic properties of the base words can be built into the word formation rules for morphological processes, token frequencies do not lend themselves for a formal representation, and, more importantly, the interactions in the mental lexicon between stored items, which are only imperfectly summarized in terms of frequency, escape such formalization altogether. Anshen and Aronoff (1988) recognize the relevance of token frequencies, but evade the problem that the rules of Item and Process models are not designed for dealing with token frequencies by stipulating that productive formations are not stored at all. This amounts to the claim that frequency is simply irrelevant for productive rules. As argued above, this claim cannot be upheld. If the role of frequency is to be incorporated within the framework of Item and Process theory, perhaps the best way to proceed is to calculate the degree of productivity \mathcal{P} for each derivational subdomain, thus enriching the formal statement of the word formation process with a description of its productivity. However, other theoretical frameworks that do not strictly separate rules from items have been developed, and it is to two such theories that we now turn.

Bybee (1985, 1988) develops a 'dynamic model of lexical representation' as an alternative for Item and Process models. In this model, morphological rules and lexical representations are not strictly separated. Instead, the rules are viewed as patterns that emerge from the intrinsic organization of the lexicon. Words in the lexicon are assumed to be linked by means of so-called semantic and phonological connections, which organize the words in the lexicon along lines of similarity. Furthermore, each word is assumed to have a level of lexical strength, an index of word frequency, that is roughly equivalent to what we have called activation level or representational strength. Word frequency is introduced into the model in order to account for historical and cross-linguistic effects of word frequency on morphology. For instance, the strong verbs of English that have regularized typically are low-frequency words. When such infrequently used irregular forms fade they are susceptible to being replaced by regular formations (Bybee 1985: 119–121).

Bybee's (1985) analysis of productivity is based on the observation that in

French or English the classes of regular verbs have high type frequencies and low (mean) token frequencies, while classes of irregular verbs have high (mean) token frequencies and low type frequencies. The high type frequencies of the regular classes is singled out as characteristic of productivity (1985: 133). On the other hand, the high token frequencies of the irregular verbs seem to suggest that these verbs have higher lexical strength than the regular verbs. However, Bybee (1985) argues that frequency affects not only the lexical strength, but also the connection strength. She assumes that low-frequency formations that are morphologically complex acquire stronger connections than high-frequency complex words. "This is the way in which the model represents the fact that low-frequency items are analyzed and understood in terms of other items, while high-frequency words, complex or not, may be autonomous, and processed unanalyzed" (Bybee 1985: 123–124). Hence, low-frequency regular complex words are concluded to acquire higher levels of lexical strength than high-frequency irregular complex items.

Word formation rules play a minimal rôle in this model. 'Less productive affixes' are claimed (1985: 128) to have no representation independent of the words in which they occur. In the case of 'extremely productive affixes', however, the affix may obtain an independent representation, just as the complex words to which they are attached. When a neologism has to be processed, the complex form can be arrived at by combination of stem and affix. Bybee (1985: 129) argues that the two ways of obtaining the correct form, by rote and by combination, are highly similar: even when a complex form does not already exist as a separate entry, it is implicitly present given the representations of base and affix, and the connections between all members of the morphological category involved. In Bybee (1988) she argues on the basis of the Rumelhart & McClelland (1986) study of past tense formation in English for the more radical position that rules and representations are completely merged. We return to this issue below.

Bybee's (1988) model differs on one subtle point with her (1985) analysis. As mentioned above, Bybee (1985) argues that token frequency and connection strength are inversely related. This is a rather counterintuitive assumption that is in conflict with the normal direction of frequency effects. This assumption is dropped in Bybee (1988), where it is claimed, rather than explained in terms of frequency-determined connection strength, that high-frequency simplex words are acquired more or less independently of other words and hence have few connections to other items, while low-frequency complex words are learned and stored in terms of the more basic words that already have lexical representations in the mental lexicon, taking on many connections with other items. Since larger numbers of connections running in parallel may accumulate to form strong connections, frequency and connection strength are again inversely correlated, but now only indirectly so. We may note that this correlation is linked to the observation that generally high-frequency words have more idiosyncratic properties than low-frequency words. In Bybee's model, this insight translates directly into the statement that low-frequency words will have more connections than high-frequency words. However, this approach leads to the strong claim that high frequency

and irregularity are two sides of the same coin. Unfortunately, as we shall see below, this claim is somewhat too strong.

Having sketched the outlines of Bybee's dynamic model, we now turn to consider her analysis of productivity in some more detail. As mentioned above, Bybee rejects the idea that token frequencies are directly relevant here. With respect to past tense formation in English, she remarks that (1988: 138)

A larger number of distinct verbs participating in the same pattern will serve to strengthen it. Note that it is type frequency rather than token frequency. A verb of high token-frequency will not serve to strengthen a schema; in fact, it appears that very high-frequency verbs have very little effect on productivity, since ... such forms seem to be processed without forming connections with other items.

To our mind, this analysis is too simplistic, for two reasons. First, as pointed out in Section 2, this emphasis on type frequency is only partially correct. Some productive processes show up with large numbers of types, but others are found with type frequencies that are lower than those of unproductive processes. Crucial to the productivity of a morphological process is that there be enough low-frequency types, even though the total number of types may be quite small, as in the case of Dutch *-sel*. A large proportion of low-frequency types is also significant in the sense that, given decay of lexical strength over time, word formation rules may well be involved in keeping the lexical representations available.

Second, as was argued above, Bybee's inverse relation between frequency and number (and strength) of connections derives from the idea that frequency and regularity are inversely proportional. Hence, Bybee's analysis of productivity in terms of token frequencies is ultimately based on the idea that productivity and regularity coincide. Unfortunately, the notions 'productivity' and 'regularity' are logically independent, since unproductive processes may be regular. Regular unproductive formations such as Dutch de-adjectival *-te* combine the property of having many connections with the property of having high token frequencies, contrary to the inverse relation between number of connections and frequency in Bybee's theory. In other words, a high-frequency word need not be irregular, and may be well integrated in the network of lexical relations. This network may include only high-frequency formations, in which case we are dealing with an unproductive process, or it may contain large numbers of low-frequency formations, in which case we have a productive process. If we conceive of connections as the pathways for mutual re-enforcement of lexical strength, we find that high-frequency productive complex formations may strengthen the lexical representations of low-frequency formations of similar morphological make-up. Of course, type frequency is not irrelevant here, since large numbers of such low-frequency formations in for instance *-ness* will strengthen a given pattern of affixation more than one or two medium frequency formations. In sum, one should not trade off token frequency against type frequency. Instead, principled means should be found to evaluate the contributions of both to the strength of patterns of affixation.

Although we have been critical of Bybee's handling of productivity, her dynamic model of lexical representation as outlined in Bybee (1988) presents a useful framework for the analysis of this morphological phenomenon. The introduction of connections between entries in the lexicon allows for the integration of frequency data and patterns of regularities in a natural way. A weak point, however, is that the model is very implicit, especially with respect to the layout of the network and the way in which the network is the rule. However, Bybee (1988) argues that the results obtained by Rumelhart and McClelland (1986) with an explicitly defined network of lexical connections shows that formal implementations of her approach can be worked out.

Rumelhart and McClelland (1986) discuss a computer simulation of past tense formation in English, using a connectionist pattern matcher. The details of the architecture of the model need not concern us here. Essential is the fact that input and output phonological units are associated by means of a network of weighted connections. This pattern matcher was repeatedly exposed to a list of 420 English verbs, 20% of which were irregular. By adjusting the connection weights after each exposure whenever the output was incorrect, the model eventually succeeded in matching present tense verbs in its input with the correct past tense forms of these verbs. This result was obtained without the use of explicit rules or schema's: lexical representations and word formation rules are completely merged.

Rumelhart and McClelland (1986) use this result to argue for sub-symbolic models of human cognition. Unfortunately, a number of substantial problems remain to be solved before their claim can be accepted (see e.g. Massaro 1988, Pinker and Prince 1988). However, the Rumelhart and McClelland (1986) pattern matcher can be used, not to argue for or against sub-symbolic modelling, but to gain some insight into the effects of token frequency on rules in dynamic models and, for instance, into the rôle of type frequency on the productivity of morphological processes.

Evidence that the rôle of type frequency in these models should not be overestimated is discussed in Baayen (1989), where I studied the way in which the Rumelhart and McClelland pattern matcher handles the affixation of the Dutch rival suffixes *-heid* and *-te*. The importance of high token-frequencies for unproductive *-te* is immediately apparent in the light of the following. We trained the Rumelhart and McClelland pattern matcher by exposing it once to a list of (3751) tokens in Dutch unproductive *-te* and productive *-heid*. The tokens were presented in random order. Each type was represented by the number of tokens with which it occurs in the database.²⁰ The model assigned the suffix *-te* correctly to 71% of the adjectival base words which take *-te*, and attaches *-heid* correctly to 98% of the base words which occur with *-heid* in the EC. Following this, we trained the model on a list with the same number of tokens, but now each type in this list had the same token frequency. The scores of correct assignment for *-te* and *-heid* obtained drop to 30% and rise to 100% respectively. Clearly, the small number of types in *-te* (roughly 40) can be generated only when the token frequencies of these types are high. This result is in line with the results obtained by Rumelhart and McClelland (1986) and Plunkett and Marchman

(1989) for the irregular past tense forms. It is to be expected since *-te* attaches to base words that cannot be distinguished phonologically from (a subset of) the base words to which *-heid* is suffixed.

We are now in the position to evaluate the effects of type frequency in this particular instance of a dynamic model. In contrast to *-heid*, *-te* attaches only to simplex base words. In order to ascertain the effect of the formations in *-heid* coined from complex base words on the performance of the model with respect to formations in *-te* and *-heid* coined from simplex adjectives, we trained the model on a list that contained simplex formations only. Again each type in this list was represented by the number of tokens with which it occurs in the EC. Surprisingly, the scores obtained (70% for *-te*, 90% for *-heid*) do not differ significantly ($p > 0.05$) from the scores obtained when the model is trained on both simplex and complex formations. In other words, the presence or absence of some four hundred second order complex types in *-heid* does not seriously affect the processing of the formations in *-heid* and *-te* coined from simplex base words. This result suggests that type frequency by itself is not the prime factor which conditions productivity. Instead, type frequency, token frequency and shared patterns of phonological similarity jointly determine the qualitative output of the model. Interestingly, the reason that the types in *-heid* derived from complex adjectives do not affect the derivation of *-te* and *-heid* coined from simplex base words resides in the fact that the final rhymes of the complex adjectival base words are in complementary distribution with the rhymes of the simplex adjectives which yield abstract nouns in *-te*. Put differently, a morphological condition on the word formation rule of *-te*, namely that it does not accept complex adjectives as input, can be re-analysed as being of a phonological nature, in that only a specific gang of phonologically similar base words allow *-te* to be attached. This insight is difficult to formulate within an Item and Process model, but follows naturally from the representational approach advocated by Bybee and Rumelhart and McClelland.

Let us summarize our findings. Item and Process models of the kinds developed by morphologists within generative or functional paradigms are especially well-suited for tracing the various qualitative restrictions on the productivity of word formation rules. The weak point of these models is that they do not lend themselves for the analysis of the rôle of token frequency in morphology. Since the primary aim of these models is to analyse the language system rather than language use, this need not be a problem. On the other hand we may note that dynamic models of the kind proposed by Bybee have the advantage that their design immediately captures the rôle of token frequency. However, theories which make extensive use of connections in the lexicon must be explicitly worked out, if such theories are to be more than metaphorical descriptions of possible mental processing mechanisms. The Rumelhart and McClelland (1986) connectionist model illustrates that a network theory can be articulated. However, the lexical representations of network models should be chosen with care, since they bear the burden of expressing the valuable insights expressed by the rules of Item and Process models. The connectionist pattern matcher is in this regard highly unsatis-

factory. More elaborate networks, analogical models along the lines proposed by Skousen (1989), or other forms of cluster analysis may provide alternative lines of inquiry, but here too the results obtained will hinge on the coding of the input. It is to be hoped that future research will lead up to models which combine the strong points of both the 'generative' and the 'dynamic' approaches to morphology.

NOTES

* The author is indebted to Geert Booij, Pieter van Reenen, Rochelle Lieber and Ariane van Santen for valuable discussion on the linguistic interpretation of the statistics developed here, and Richard Gill and Rezo Chitašvili for their aid in coming to grips with the mathematics of frequency distributions. Finally, I have been able to benefit from discussions with Uli Frauenfelder, Robert Schreuder and Willem Levelt on the psycholinguistic aspects of productivity. All errors and follies in this paper remain the responsibility of the author.

¹ This list should be extended with semantic conditioning factors. For instance, Zimmer (1964) points out that English *un-* tends not to attach to any base which is semantically negative (**unbad*, **unsick*), and Rainer (1988) calls attention to the fact that affixes deriving abstract quality nouns can only be attached to qualitative, but not to relational adjectives (compare *goodness* with **woodness*).

² For instance, French purism has been argued by Martinet (1969) and Zwanenburg (1971) to have a negative influence on the degree of productivity of morphological processes in French.

³ For instance, in advertising the trend is to ask for a *wetenschappelijk medewerker (M/V)* (a research assistant (male/female) rather than to mention explicitly that both *medewerkers* and *medewerksters* can apply.

⁴ Other statistical models available for the analysis of word frequency distributions can be found in Sichel (1975) and Carroll (1967).

⁵ Note that the values of \mathcal{P} cannot be directly compared across the two languages, as the frame samples from which the item samples have been collected are vastly different in size, 18 000 000 for the English corpus, 600 000 for the Dutch corpus. This has the distorting effect that *-heid* seems more productive than *-ness*.

⁶ It can be shown (Baayen 1989) that the difference in growth rates between simplex nouns and *-te* is not significant at the 5% level, and that the reverse holds for *-heid*.

⁷ For a detailed discussion see Baayen and Lieber (1991).

⁸ Baayen (1989) discusses these assumptions in detail. We may note that the assumption of independence is not met for running text. The rules of syntax, together with the requirements of textual coherence and cohesion cause the words of a text to be arranged in patterns that are far from random. However, our item samples do not contain all the words of some text or part of a text, hence, many factors which give rise to interdependence are irrelevant. Although it can be shown that even in the item samples the tokens do not occur independently, the effect of interdependence on the predictions of the model is minimal. In fact, the results obtained with the empirical item samples are replicated when special measures are taken to ensure that the formations in the item sample are strictly independent.

⁹ The extended Zipf-Mandelbrot law is obtained when $\alpha = \beta = 1$ and γ free.

¹⁰ Although the value of \mathcal{P} suggests that *-erd* is very productive the very small number of tokens involved requires caution in evaluating \mathcal{P} . The problem is that for such low values of N the mathematics underlying the interpretation of \mathcal{P} do not apply. On the other hand, truly unproductive processes tend to be represented by few but high frequency types. To give an example from inflection, the Dutch unproductive plural ending *-eren* (*kind-eren*, 'children') shows up with 9 types in the EC. However, the number of tokens equals 459, and $\mathcal{P} = 0.002$.

¹¹ When comparing the number of types of morphological processes across corpora, V loses its interpretation of extent of use, an interpretation which hinges on the fact that the

morphological processes being compared are sampled within the same overall frame sample of size F . Hence, comparisons in terms of extent of use (and degree of productivity) across corpora are legitimate only when these corpora are of approximately equal size.

¹² For a detailed discussion of the productivity of these and other affixes in English see Baayen and Lieber, MS.

¹³ Highly unproductive word formation processes, such as the process underlying action nouns with vocalic alternation in Dutch, are characterized by grouped frequency distributions for which the mode does not equal unity. For such frequency distributions, Zipf's law and its extensions are invalid.

¹⁴ The source of this paradoxical state of affairs may reside in the stochastic process which yields (simplex) words. Nusbaum (1985) discusses results obtained with a computer simulation of a Zipf-like stochastic process which produced a set of words for which the size of the similarity neighbourhoods of high and low frequency words were found to differ consistently. As in natural language, high-frequency strings turned out to have large numbers of (high-frequency) neighbours while low-frequency strings were found to have significantly smaller sets of (low-frequency) neighbours. This relationship between frequency, length and the size of the similarity neighbourhood may underlie the fact that subjects have been found to be able to consistently estimate the frequency of nonwords (Eukel 1980).

¹⁵ Bradley (1979) reports that the cumulative root frequency of the transparent productive affixes *-er*, *-ment*, *-ness* appears to be the sole determinant of response latencies. For less transparent *-ion*, neither word frequency nor cumulative root frequency were found to be reliable predictors of response latencies. However, Cole *et al.* (1989) and Burani and Caramazza (1987) obtained different experimental results indicating that word frequency and cumulative root frequency jointly determine response latencies.

¹⁶ This result also follows immediately from Zipf's frequency-length relation, when length is expressed in terms of the number of morphemes. This relationship expresses the observation that frequent words tend to be short. Hence, the hapaxes will typically contain the highest percentage of words with large numbers of morphemes. Conversely, words with large numbers of hapaxes will manifest the highest degree of productivity.

¹⁷ The figure n_0 represents the number of adjectives that do not occur in the EC but which serve as base words for attested formations in *-heid* and *-te*. The difference in the number of such types, 2 for *-te* and 74 for *-heid* suggests that a significant difference relating to the productivity of these affixes is involved. However, the difference in the ratios n_0/V is not significant ($p > 0.05$, using the exact Fisher test of independence (Sacks 1982: 370–373)).

¹⁸ In fact, the type-token ratios presented by Anshen and Aronoff (1988) have been calculated in such a way that the mean token frequency of formations in *-ness* seems to be negligible indeed. For instance, their Table 4 on p. 645 lists a mean token frequency of 0.49 for words in *-iveness* and 9.57 for words in *-ivity*. However, these ratios are obtained by taking into account both the types which occur in the Brown corpus and the types which do not occur in this corpus but which are listed in Walker (1936). This mixing of frequency data from corpus and dictionary is statistically illegitimate — it is entirely unclear on what kind of sample space our probability measure has to be defined — and has the effect of exaggerating the difference in token frequency between *-ness* and *-ity*. See Baayen and Lieber (1991) for further details.

¹⁹ See Baayen and Lieber (1991) for a more detailed discussion of the productivity of *-ness* and *-ity* across various subdomains.

²⁰ In this case the data were obtained from the corpora of Uit den Boogaart (1975) and de Jong (1977).

REFERENCES

- Anshen, F. and M. Aronoff. 1988. "Producing Morphologically Complex Words". *Linguistics* 26, 641–655.
 Aronoff, M. 1976. *World Formation in Generative Grammar*. Cambridge, Mass.: MIT Press.
 Aronoff, M. 1982. "Potential Words, Actual Words, Productivity and Frequency". In *Preprints*

of the Plenary Session Papers. The XIIIth International Congress of Linguists, Tokyo, 141–148.

- Baayen, R. H. 1989. *A Corpus-Based Approach to Morphological Productivity. Statistical Analysis and Psycholinguistic Interpretation*. Dissertation. Vrije Universiteit, Amsterdam.
 Baayen, R. H. and R. Lieber. 1991. "Productivity and English Derivation: A Corpus Based Study". *Linguistics* 29, 801–843.
 Balota, D. A. and J. I. Chumbley. 1984. "Are Lexical Decisions a Good Measure of Lexical Access? The Role of Word Frequency in the Neglected Decision Stage". *Journal of Experimental Psychology: Human Perception and Performance* 10, 340–357.
 Balota, D. A. and J. I. Chumbley. 1985. "The Locus of Word Frequency Effects in the Pronunciation Task: Lexical Access and/or Production?". *Journal of Memory and Language* 24, 89–106.
 Bolinger, D. L. 1948. "On Defining the Morpheme". In D. L. Bolinger (ed.), *Forms of English. Accent, Morpheme, Order*, Cambridge, Mass.: Harvard University Press, 183–189.
 Booij, G. E. 1977. *Dutch Morphology. A Study of Word Formation in Generative Grammar*. Dordrecht: Foris.
 Bradley, D. C. 1979. "Lexical Representation of Derivational Relation". In M. Aronoff and M. L. Kean (eds.), *Juncture*. Saratoga: Anma Libri, 37–55.
 Bradley, D. C. and K. I. Forster. 1987. "A Reader's View of Listening". *Cognition* 25, 103–134.
 Brekle, H. E. and D. Kastovsky (eds.). 1977. *Perspektiven der Wortbildungsforschung*. Bonn: Bouvier Verlag.
 Broadbent, D. E. 1967. "Word Frequency Effect and Response Bias". *Psychological Review* 74, 1–15.
 Broadbent, D. E. and M. H. P. Broadbent. 1975. "Some Further Data Concerning the Word Frequency Effect". *Journal of Experimental Psychology: General* 104, 297–308.
 Burani, C. and A. Caramazza. 1987. "Representation and Processing of Derived Words". *Report 25, Cognitive Neuropsychology Laboratory*, The Johns Hopkins University.
 Burgschmidt, E. 1977. "Strukturierung, Norm und Produktivität in der Wortbildung". In H. E. Brekle and D. Kastovsky (eds.), 39–47.
 Butterworth, B. (ed.). 1983. *Language Production (Vol. II): Development, Writing and Other Language Processes*. London: Academic Press.
 Butterworth, B. 1983. "Lexical Representation". In B. Butterworth (ed.), 257–294.
 Bybee, J. L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: Benjamins.
 Bybee, J. L. 1988. "Morphology as Lexical Organization". In M. Hammond and M. Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*. London: Academic Press, 119–141.
 Caramazza, A., A. Laudanna, and C. Romani. 1988. "Lexical Access and Inflectional Morphology". *Cognition* 28, 297–332.
 Carroll, J. B. 1967. "On Sampling from a Lognormal Model of Word Frequency Distribution". In H. Kučera and W. N. Francis, 406–424.
 Chitašvili, R. J. and E. V. Khmaladze. 1989. "Statistical Analysis of Large Number of Rare Events and Related Problems". *Transactions of the Tbilisi Mathematical Institute* 92, 196–245.
 Cole, P., C. Beauvillain, and J. Segui. 1989. "On the Representation and Processing of Prefixed and Suffixes Derived Words: A Differential Frequency Effect". *Journal of Memory and Language* 28, 1–13.
 Coseriu, E. 1970. "System, Norm und Rede". In id., *Sprache, Strukturen und Funktionen. XII Aufsätze*. Tübingen: Narr, 193–212.
 Coseriu, E. 1975. "System, Norm und Rede". In id., *Sprachtheorie und allgemeine Sprachwissenschaft*. München: Fink, 11–101.
 Downing, P. 1977. "On the Creation and Use of English Compound Nouns". *Language* 53, 810–842.
 Efron, B. and R. Thisted. 1976. "Estimating the Number of Unseen Species: How Many Words did Shakespeare Know?". *Biometrika* 63, 435–447.

- Eukel, B. 1980. "A Phonotactic Basis for Word Frequency Effects: Implications for Automatic Speech Recognition". *Journal of the Acoustical Society of America* 68, S33.
- Forster, K. I. 1976. "Accessing the Mental Lexicon". In R. J. Wales and E. Walker (eds.), 257–287.
- Forster, K. I. and S. M. Chambers. 1973. "Lexical Access and Naming Time". *Journal of Verbal Learning and Verbal Behavior* 12, 627–635.
- Gardner, M. K. and E. Z. Rothkopf, R. Lapan, and T. Lafferty. 1987. "The Word Frequency Effect in Lexical Decision: Finding a Frequency-Based Component". *Memory and Cognition* 15, 24–28.
- Geerts, G., W. Haeseryn, J. de Rooij, and M. C. van den Toorn (eds.). 1984. *Algemene Nederlandse Spraakkunst*. Groningen, Wolters-Noordhoff.
- Good, I. J. and G. H. Toulmin. 1956. "The Number of New Species and the Increase in Population Coverage, When a Sample is Increased". *Biometrika* 43, 45–63.
- Grainger, J., J. K. O'Regan, A. M. Jacobs, and J. Segui. 1989. "On the Role of Competing Word Units in Visual Word Recognition: The Neighborhood Frequency Effect". *Perception & Psychophysics* 45, 189–195.
- Guiter, H. and M. V. Arapov (eds.). 1983. *Studies on Zipf's Law*. Bochum: Brockmeyer.
- Harwood, F. W. and A. M. Wright. 1956. "Statistical Study of English Word Formation". *Language* 32, 260–273.
- Hasher, L. and R. T. Zacks. 1984. "Automatic Processing of Fundamental Information. The Case of Frequency of Occurrence". *American Psychologist* 39, 1372–1388.
- Herdan, G. 1960. *Type-Token Mathematics*. The Hague: Mouton.
- Herdan, G. 1964. *Quantitative Linguistics*. London: Butterworths.
- Hockett, C. 1954. "Two Models of Grammatical Description". *Word* 10, 210–231.
- Jakobson, R. (ed.). 1962. *Structure of Language and its Mathematical Aspects. Proceedings of Symposia in Applied Mathematics*, Vol. XII, Providence, Rhode Island, American Mathematical Society.
- Jarvella, R. J., G. Sandström and R. Schreuder. 1987. "Morphological Constraints on Word Recognition". In A. Allport, D. G. MacKay, W. Prinz and E. Scheerer (eds.), *Language Perception and Production: Relationships Between Listening, Speaking, Reading, and Writing*. London: Academic Press, 245–265.
- Jong, E. D. de. 1979. *Spreektaal. Woordfrequenties in Gesproken Nederlands*. Utrecht: Oosthoek, Scheltema en Holkema.
- Kalnin, V. M. 1965. "Functionals Related to the Poisson Distribution, and Statistical Structure of a Text". *Proceedings of the Steklov Institute of Mathematics* 79; translation in: J. V. Fennik (ed.): *Articles on Mathematical Statistics and the Theory of Probability*. American Mathematical Society, Providence, Rhode Island, 202–220.
- Khmaladze, E. 1987. "The Statistical Analysis of a Large Number of Rare Events". *Report MS-R8804*, Dept. of Mathematical Statistics, Center for Mathematics and Computer Science, Amsterdam.
- Kučera, H. and W. N. Francis. 1967. *Computational Analysis of Present-Day American English*. Providence: Brown University Press.
- Landauer, T. K. and L. A. Streeter. 1973. "Structural Differences Between Common and Rare Words: Failure of Equivalence Assumptions for Theories of Word Recognition". *Journal of Verbal Learning and Verbal Behavior* 12, 119–131.
- Laudanna, A. and C. Burani. 1985. "Address Mechanisms to Decomposed Lexical Entries". *Linguistics* 23, 775–792.
- Lima, S. D. 1987. "Morphological Analysis in Sentence Reading". *Journal of Memory and Language* 26, 84–99.
- Luce, P. A. 1986. "Neighborhoods of Words in the Mental Lexicon". *Research on Speech Perception, Technical Report # 6*, Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington.
- MacWhinney, B. 1978. *The Acquisition of Morphophonology*. Child Development Publication. Chicago: University of Chicago Press.
- Mandelbrot, B. 1962. "On the Theory of Word Frequencies and on Related Markovian Models of Discourse". In Jakobson (ed.), 190–219.

- Manelis, L. and D. A. Tharp. 1977. "The Processing of Affixed Words". *Memory and Cognition* 5, 690–695.
- Marle, J. van. 1985. *On the Paradigmatic Dimensions of Morphological Creativity*. Dordrecht: Foris.
- Marle, J. van. 1988. "Betekenis als Factor bij Productiviteitsverandering". *Spektator* 17, 341–359.
- Marslen-Wilson, W. D. 1987. "Functional Parallelism in Spoken Word-Recognition". *Cognition* 25, 71–102.
- Martinet, A. 1969. "Les puristes contre la langue". In id., *Le Français sans Fard*, Paris, 25–32.
- Massaro, D. W. 1988. "Some Criticisms of Connectionist Models of Human Performance". *Journal of Memory and Cognition* 27, 213–234.
- McClelland, J. L. and D. E. Rumelhart (eds.). 1986. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*. Vol. 2, *Psychological and Biological Models*, Cambridge, Mass.: MIT Press.
- McRae, K., D. Jared, and M. S. Seidenberg. 1990. "On the Roles of Frequency and Lexical Access in Word Naming". *Journal of Memory and Language* 29, 43–65.
- Meys, W. 1985. "Morfologische Regels en het Mentale Lexicon". *Glott* 8, 183–206.
- Monsell, S. 1985. "Repetition and the Lexicon". In A. W. Ellis (ed.), *Progress in the Psychology of Language* (Vol. 2), London: Erlbaum.
- Monsell, S., M. C. Doyle, and P. N. Haggard. 1989. "Effects of Frequency on Visual Word Recognition Tasks: Where are They?". *Journal of Experimental Psychology: General* 118, 43–71.
- Morton, J. 1968. "A Retest of the Response-Bias Explanation of the Word Frequency Effect". *The British Journal of Mathematical and Statistical Psychology* 21, 21–33.
- Muller, Ch. 1979a. "Peut-on estimer l'étendue d'un lexique?". In Muller 1979c, 399–425.
- Muller, Ch. 1979b. "Du nouveau sur les distributions lexicales: la formule de Waring-Herdan". In Muller (1979c), 177–195.
- Muller, Ch. 1979c. *Langue Française et Linguistique Quantitative*. Genève: Slatkine.
- Nusbaum, H. C. 1985. "A Stochastic Account of the Relationship between Lexical Density and Word Frequency". *Research on Speech Perception, Progress Report # 11*, Indiana University.
- Orlov, J. K. 1983a. "Dynamik der Häufigkeitsstrukturen". In H. Guiter and M. V. Arapov (eds.), 116–153.
- Orlov, J. K. 1983b. "Ein Model der Häufigkeitsstruktur des Vokabulars". In H. Guiter and M. V. Arapov (eds.), 154–233.
- Orlov, J. K. and R. Y. Chitašvili. 1982a. "On the Distribution of Frequency Spectrum in Small Samples from Populations with a Large Number of Events". *Bulletin of the Academy of Sciences, Georgia* 108, 297–300.
- Orlov, J. K. and R. Y. Chitašvili. 1982b. "On Some Problems of Statistical Estimation in Relatively Small Samples". *Bulletin of the Academy of Sciences, Georgia* 108, 513–516.
- Orlov, J. K. and R. Y. Chitašvili. 1983a. "On the Statistical Interpretation of Zipf's Law". *Bulletin of the Academy of Sciences, Georgia* 109, 505–508.
- Orlov, J. K. and R. Y. Chitašvili. 1983b. "Generalized Z-Distribution Generating the Well-Known 'Rank-Distributions'". *Bulletin of the Academy of Sciences, Georgia* 110, 269–272.
- Pinker, S. and A. Prince. 1988. "On Language and Connectionism". *Cognition* 28, 73–193.
- Pisoni, D. B., H. C. Nusbaum, P. A. Luce, and L. M. Slowiaczek. 1985. "Speech Perception, Word Recognition and the Structure of the Lexicon". *Speech Communication* 4, 75–95.
- Plunkett, K. and V. Marchman. 1989. "Pattern Association in a Back Propagation Network: Implications for Child Language Acquisition". *Technical Report # 8902*, Center for Research in Language, University of California, San Diego.
- Rainer, F. 1988. "Towards a Theory of Blocking: the Case of Italian and German Quality Nouns". *Yearbook of Morphology* 1, 155–185.
- Rayner, K. and S. A. Duffy. 1986. "Lexical Complexity and Fixation Times in Reading: Effects of Word Frequency, Verb Complexity, and Lexical Ambiguity". *Memory and Cognition* 14, 191–201.

- Renouf, A. 1987. "Corpus Development". In J. M. Sinclair (ed.), *Looking Up: An Account of the Cobuild Project in Lexical Computing*. Collins.
- Riddle, E. A. 1984. "A Historical Perspective on the Productivity of the Suffixes *-ness* and *-ity*". *Conference on Historical Semantics and Word Formation, Blazjewko, Poland*, 28–31.
- Rubenstein, H. and I. Pollock. 1963. "Word Predictability and Intelligibility". *Journal of Verbal Learning and Verbal Behavior* 2, 147–158.
- Rumelhart, D. E. and J. L. McClelland. 1986. "On Learning the Past Tenses of English Verbs". In J. L. McClelland and D. E. Rumelhart (eds.), 216–271.
- Rumelhart, D. E. and J. L. McClelland (eds.). 1986. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*. Vol. 1 *Foundations*, Cambridge, Mass.: MIT Press.
- Sachs, L. 1982. *Applied Statistics. A Handbook of Techniques*. New York: Springer Verlag.
- Santen, A. van and J. W. de Vries. 1981. "Vrouwelijke Persoonsnamen op *-ster*". *Forum der Letteren* 22, 115–125.
- Scarborough, D. L., C. Cortese, and H. S. Scarborough. 1977. "Frequency and Repetition Effects in Lexical Memory". *Journal of Experimental Psychology: Human Perception and Performance* 3, 1–17.
- Schreuder, R. 1990. "Lexical Processing of Verbs with Separable Particles". *Yearbook of Morphology* 3, 65–79.
- Schreuder, R., M. Grendel, N. Poulisse, A. Roelofs, and M. van der Voort. 1990. "Lexical Processing, Morphological Complexity and Reading". In D. A. Balota, G. B. Flores d'Arcais and K. Rayner (eds.), *Comprehension Processes in Reading*. London: Lawrence Erlbaum Associates.
- Schultink, H. 1961. "Productiviteit als Morfologisch Fenomeen". *Forum der Letteren* 2, 110–125.
- Schultink, H. 1962. *De Morfologische Valentie van het Ongelede Adjectief in Modern Nederlands*. Den Haag: van Goor Zonen.
- Sichel, H. S. 1975. "On a Distribution Law for Word Frequencies". *Journal of the American Statistical Association* 70, 542–547.
- Simon, H. A. 1955. "On a Class of Skew Distribution Function". *Biometrika* 42, 435–440.
- Simon, H. A. 1960. "Some Further Notes on a Class of Skew Distribution Functions". *Information and Control* 3, 80–88.
- Skousen, R. 1989. *Analogical Modeling of Language*. Dordrecht: Kluwer.
- Stanners, R. F. and J. J. Neiser, W. P. Hernon, and R. Hall. 1979. "Memory Representation for Morphological Related Words". *Journal of Verbal Learning and Verbal Behavior* 18, 399–412.
- Stemberger, J. P. and B. MacWhinney. 1986. "Frequency and the Lexical Storage of Regularly Inflected Forms". *Memory and Cognition* 14, 17–26.
- Stemberger, J. P. and B. MacWhinney. 1988. "Are Inflected Forms Stored in the Lexicon?". In M. Hammond and M. Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*. London: Academic Press, 101–116.
- Taft, M. 1979. "Recognition of Affixed Words and the Word Frequency Effect". *Memory and Cognition* 7, 263–272.
- Taft, M. 1988. "A Morphological-decomposition Model of Lexical Representation". *Linguistics* 26, 657–667.
- Taft, M. and K. I. Forster. 1975. "Lexical Storage and the Retrieval of Prefixed Words". *Journal of Verbal Learning and Verbal Behavior* 14, 638–647.
- Tyler, L. K., W. Marslen-Wilson, and P. Hanney. 1988. "Continuous and Discontinuous Access in Spoken Word-Recognition: The Role of Derivational Prefixes". *Journal of Memory and Language* 27, 368–381.
- Uhlenbeck, E. M. 1977. "The Concepts of Productivity and Potentiality in Morphological Description and their Psychological Reality". *Salzburger Beiträge zur Linguistik* 4, 379–391.
- Uit den Boogaart, P. C. (ed.). 1975. *Woordfrequenties in Gesproken en Geschreven Nederlands*. Utrecht: Oosthoek, Scheltema & Holkema.

- Wales, R. J. and E. Walker (eds.). 1976. *New Approaches to Language Mechanisms. A Collection of Psycholinguistic Studies*. Amsterdam: North-Holland.
- Walker, J. 1936. *Walker's Rhyming Dictionary*. New York: Dutton.
- Whaley, C. P. 1978. "Word-Nonword Classification Time". *Journal of Verbal Learning and Verbal Behavior* 17, 143–154.
- Zimmer, K. E. 1964. *Affixal Negation in English and Other Languages*. Supplement to *Word* 20.
- Zipf, G. K. 1935. *The Psycho-Biology of Language*. Boston: Houghton Mifflin.
- Zwanenburg, W. 1971. *Franse Afleidingsmanoeuvres*. Leiden: Universitaire Pers.

Max-Planck-Institut für Psycholinguistik,
Wundtlaan 1
6525 XD Nijmegen, The Netherlands.
e-mail: baayen@mpi.nl