Matthews, P. H. 1972. *Inflectional Morphology*. Cambridge: Cambridge University Press.
McIntosh, M. 1984. *Fulfulde Syntax and Verbal Morphology*. Boston: KPI in association with University of Port Harcourt Press.
Muysken, P. 1986. "Approaches to Affix Order". *Linguistics* 24, 629—643.
Pesetsky, D. 1985. "Morphology and Logical Form". *Linguistic Inquiry* 16, 193—246.
Platt, D. 1981. "Old Provençal — The Balance between Regularity and Irregularity in Verbal Inflection". In T. Thomas-Flinders (ed.), 41—71.
Pullum, G. K. and A. M. Zwicky. To appear. "A Misconceived Approach to Morphology". *Proceedings of the Tenth West Coast Conference on Formal Linguistics*.
Sadock, J. M. 1991. *Autolexical Syntax: A Theory of Parallel Grammatical Relations*. Chicago: University of Chicago Press.
Scalise, S. 1988. "Inflection and derivation". *Linguistics* 26, 561—581.
Schadeberg, T. C. 1984. *A Sketch of Swahili Morphology*. (2nd revised ed.). Dordrecht: Foris.
Schadeberg, T. C. 1989. "The Three Relative Constructions in Swahili". In M.-F. Rombi (ed.), *Le Swahili et ses limites*. Paris: Editions Recherche sur les Civilisations, 33—39.
Schultink, H. 1988. "Morphological Heads: Evidence from Swahili". In G. Booij and J. van Marle (eds.), *Yearbook of Morphology 1988*. Dordrecht: Foris, 247—258.
Selkirk, E. O. 1982. *The Syntax of Words*. Cambridge Mass.: MIT Press.
Sproat, R. W. 1985. *On Deriving the Lexicon*. Doctoral dissertation, MIT.
Stump, G. T. 1989. "A Note on Breton Pluralization and the Elsewhere Condition". *Natural Language and Linguistic Theory* 7, 261—273.
Stump, G. T. 1991a. "A Paradigm-based Theory of Morphosemantic Mismatches". *Language* 67, 675—725.
Stump, G. T. 1991b. "On the Theoretical Status of Position Class Restrictions on Inflectional Affixes". In G. Booij and J. van Marle (eds.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer, 211—241.
Stump, G. T. To appear. "Two Types of Mismatch between Morphology and Semantics". In E. Schiller (ed.), *Developments in Autolexical Syntax*. The Hague: Mouton.
Thomas-Flinders, T. 1981. "Initial Lenition in Celtic: Evidence for Inflection as a Phonological Process". In T. Thomas-Flinders (ed.), 72—83.
Thomas-Flinders, T. (ed.). 1981. *Inflectional Morphology: Introduction to the Extended Word-and-Paradigm Theory (UCLA Occasional Papers # 4: Working Papers in Morphology)*. Los Angeles: UCLA Department of Linguistics.
Thomas-Flinders, T. 1983. *Morphological Structures*. Doctoral dissertation, UCLA.
Williams, E. 1981. "On the Notions 'Lexically Related' and 'Head of a Word'". *Linguistic Inquiry* 12, 245—274.
Zwicky, A. M. 1985. "How to Describe Inflection". *Berkeley Linguistics Society* 11, 372—386.
Zwicky, A. M. 1987. "Phonological and Morphological Rule Interactions in Highly Modular Grammars". *ESCOL '86*, 523—532.
Zwicky, A. M. 1989. "Inflectional Morphology as a (Sub)component of Grammar". In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer and J. R. Rennison (eds.), *Contemporary Morphology*. Berlin: Walter de Gruyter, 217—236.
Zwicky, A. M. 1992. "Some Choices in the Theory of Morphology". In R. D. Levine (ed.), *Formal Grammar: Theory and Implementation*. Oxford: Oxford University Press.

Department of English,
University of Kentucky,
Lexington, KY 40506, U.S.A.
e-mail: engl01 @ukcc.uky.edu

Discussion

On frequency, transparency and productivity*

HARALD BAAYEN

0. INTRODUCTION

Morphological productivity arises through a complex interaction between language structure, processing complexity and social convention. In Baayen (1989, 1991) I developed two complementary methods for the quantitative evaluation of this phenomenon. The first, and computationally most convenient method is to assess what Baayen and Lieber (1991) call the global productivity of a word formation process in terms of the number of different types V and the probability of encountering new types, $\mathscr{P} = n_1/N$, where n_1 denotes the number of types with the required affix that occur only once (the so-called hapax legomena) and N the total number of tokens with this affix in some corpus. The number of types V was interpreted as a measure of the extent of use, $\mathscr{P}$ as a measure of the degree of productivity. Baayen and Lieber (1991) applied this method in detail to a representative sample of English derivational processes. They observed that the quantitative results obtained accorded reasonably well with intuitive judgements of productivity. The second, equally valid but computationally more costly method for evaluating morphological productivity proceeds in terms of estimates of the numbers of possible types S calculated on the basis of the frequency spectra of morphological categories.

The correlation between productivity and frequency tapped into by these productivity measures is relevant to the psychological status of productive and unproductive rules. In Baayen (1991) I argued that the dominance of high frequency types in the frequency spectra of unproductive word formation processes ties in nicely with the fact that the processing of unproductive formations depends exclusively on memory storage. Since no rule is available, the only way in which the phonological and semantic properties of such formations can be accessed is by retrieval from memory. High frequencies of use ensure that these words remain available to the language user. Conversely, the large numbers of extremely low frequency types in the frequency spectra of typically productive processes suggest that whole-word storage is less relevant here. In fact, since the memory traces of low-frequency complex words are weak at best, the likelihood of morphological rules being involved in the production and perception of these words is high. If rules are indeed involved in the processing of substantial proportions of the types observed for productive affixes, this in turn suggests that the ease with which novel forms are processed is a function of the extent to which rules are involved in the processing of item-familiar (Meijs 1985) words.

The aim of the present paper is to examine the relation between morphological productivity and word frequency in some more detail in the light of a series of interesting issues raised by van Marle (1991) and Frauenfelder and Schreuder (1991).

Van Marle (1991) focuses on the empirical adequacy of the productivity statistic $\mathscr{P}$. He raises the interesting question how the productivity, measured in terms of $\mathscr{P}$, of an evidently productive process such as nominal compounding can turn out to be lower than that of a marginally productive suffix such as Dutch *-erd*, used to coin slightly pejorative personal names (*gekkerd*, 'fool', from *gek*, 'crazy'). According to van Marle, such counterintuitive results are due to a too simplistic conception of the relation between productivity and frequency.

Frauenfelder and Schreuder (1991), developing a dual route model for lexical access that aims at explaining the apparent ease with which novel forms are processed, stress the importance of phonological and semantic transparency for productivity. They argue that the approach outlined in Baayen (1991) is incomplete in that only the factor determining the speed of the memory-based access route (word frequency) is identified while the factors determining the speed of the parsing route (phonological and semantic transparency) are left unspecified.

Since the issues raised by van Marle and Frauenfelder and Schreuder are central to the understanding of morphological productivity and the way in which productivity is reflected in frequencies of use, the empirical validity of $\mathscr{P}$ is re-considered in Section 1. The relationship between transparency, frequency and productivity is re-examined in Section 2. A new productivity measure is developed by means of which the interpretational difficulties underlying the problems discussed by van Marle (1991) can be avoided. A psycholinguistic rationale for this measure is proposed that builds on a modified version of the morphological race model (henceforth MRM) developed by Frauenfelder and Schreuder (1991).

1. MEASURING PRODUCTIVITY

The main issue raised by van Marle (1991) concerns the empirical validity of the productivity statistic $\mathscr{P}$. Van Marle points out that it is impossible to obtain intuitively satisfying productivity rankings on the basis of $\mathscr{P}$. To see this, consider Table 1, which lists some Dutch morphological categories ranked intuitively according to decreasing productivity. Clearly, a ranking on the basis of $\mathscr{P}$ would require a radical re-ordering of the morphological categories listed here. This suggests that the validity of $\mathscr{P}$ as a quantitative measure of productivity seems doubtful at best. According to van Marle, the apparent failure of $\mathscr{P}$ as a productivity measure should be traced to a doubtful status assigned to hapaxes, to a too simplistic conception of the relationship between productivity and frequency and to a neglect of morphological structure.

Table 1. The measures V and $\mathscr{P}$ for a selection of Dutch morphological categories listed in order of decreasing productivity.

	V	$\mathscr{P}$
nominal compounds	4277	0.225
-tje	1031	0.253
-ing	942	0.038
-heid	466	0.114
-er	460	0.076
-ster	30	0.231
-erd	6	0.444
-sel	44	0.080
-eren	9	0.002

These claims are considered in some detail in Section 1.2. It will be shown that they can only be maintained if $\mathscr{P}$ were indeed intended to capture the *global* productivity (Baayen and Lieber 1991) of word formation processes. Since global productivity is to be evaluated in terms of both $\mathscr{P}$ *and* V, van Marle's claims lose much of their force. This is not to say, however, that the evaluation of global productivity in terms of $\mathscr{P}$ and V is properly defined. This is a serious drawback of an analysis based on $\mathscr{P}$ and V. Section 1.3 introduces a new productivity measure that by itself quantifies global productivity. First, however, it is convenient to make explicit two complementary ways in which the notion 'productivity' can be understood from a quantitative perspective.

1.1. *Aspects of productivity*

There is a broad consensus that productivity concerns the property of morphological processes to give rise to *new* words. From a quantitative point of view, there are two complementary ways in which this pretheoretical notion of productivity can be made more precise. On the one hand, we may focus on the *probability of encountering new types*. On the other hand, the notion productivity can also be understood to refer to the *total number of potential words* that can be coined on the basis of the word formation rules of the language. If the focus is to be on the number of potential words, the appropriate statistic is $\hat{S}$, the population number of types of a particular morphological constituency as estimated on the basis of the frequency spectrum of the morphological category (Baayen 1989, 1992). When we calculate $\hat{S}$ for those categories of Table 1 for which the number of observations is not too small, a satisfactory productivity ranking is obtained, as shown in Table 2.[1] The categories in Table 2 are listed according to decreasing $\hat{S}$ (column 3), where $\hat{S}$ is perhaps best interpreted as the probable

Table 2. The number of types V, the estimated number of potential types $\hat{S}$, the ratio $\mathscr{I} = \hat{S}/V$ and the number of new types $\hat{S} - V$ for a selection of Dutch morphological categories.

	V	$\hat{S}$	$\mathscr{I}$	$\hat{S} - V$
-tje	1031	$1.24\ 10^9$	$1.20\ 10^6$	$1.24\ 10^9$
nominal compounds	4277	15014	3.51	10737
-heid	466	3888	8.34	3422
-ing	942	1772	1.88	830
-er	460	1620	3.52	1160
-ster	30	382	12.73	352
-sel	44	104	2.36	60

number of types (Baayen 1989; van Santen 1992).[2] The only changes with respect to the intuition-based ranking of Table 1 concern the diminutives in *-tje* and the action nouns in *-ing*. The switch between *-ing* and *-heid* is probably due to the greater semantic transparency of de-adjectival abstract nouns in *-heid* in combination with the restricted nature of the derivational input domain of *-ing* (see Baayen 1989, 1990b). Turning to the diminutives in *-tje*, we find that $\hat{S}$ represents the number of possible rather than the number of probable types. Of all categories listed in Table 2, it is only *-tje* for which the theoretically infinite number of possible types is indeed approximated. At the same time the number of actually observed compounds is larger than that of diminutives in *-tje*. If not an artifact of the statistical method used here (see Note 1), the extremely high value of $\hat{S}$ might perhaps be traced to the fact that nominal compounding feeds diminutivization to a far greater extent than diminutivization feeds compounding. Type counts in the Uit den Boogaart (1975) corpus (henceforth UdB) show that diminutivization feeds nominal compounding for maximally 6% (236/4277) of the types, while nominal compounding feeds diminutivization for minimally 18% (187/1031) of the types ($p < 0.001$), the structure of semantically ambiguous N+N+diminutive formations having been resolved in favor of the N+[N+diminutive] analysis. If this asymmetry remains valid for ever larger corpora, it might allow diminutivization to ultimately yield a larger number of types thanks to the ease with which *-tje* attaches to compounds and the relative reluctance with which diminutives appear within compounds. A probably more important factor is that diminutivization is a semantically well-defined operation reminiscent of the semantic transparency of inflection that is applicable to any noun (with the exception of abstract nouns), while compounding requires some a priori or situational plausibility of bringing nouns together into an interpretable whole. With diminutivization as it were on the borderline between inflection and derivation — the CELEX lexical database even takes diminutivization to be an inflectional process — it is not

surprising to find the estimated population number of diminutives to approach the unlimited potentiality of inflection, surpassing even compounding.

It is possible to argue that a productivity ranking based on $\hat{S}$ does not give enough weight to the special status of *new* words. For instance, it might be argued that productivity rankings should be based on $\hat{S} - V$ rather than on $\hat{S}$. Table 2 (column 5) shows that this does not change the relative ordering of the morphological categories considered here, apart from the switch in the order of *-er* and *-ing*. Alternatively, we may shift our focus from the total number of possible types to the *extent* to which a morphological process may give rise to new types. That is, we may attempt to consider the total number of possible types *in relation to the number of actual types* (see e.g. Aronoff's (1976) index of productivity V/S) instead of the absolute number of possible types. The shift from an absolute to a relative quantification of productivity substantially changes the original ranking. For instance, when we want to investigate the extent to which the number of potential types exceeds the number of observed, actual types, we might consider a ranking based on $\mathscr{I} = \hat{S}/V$ (Baayen 1989). This results in a substantially different ordering. The suffix *-ster*, used to coin female personal nouns, is now ranked above *-er*, its unmarked counterpart. Similarly, *-heid* is now ranked above the nominal compounds. In contrast to rankings based on $\hat{S}$ or $\hat{S} - V$, an $\mathscr{I}$-based ranking quantifies the extent to which the potential of word formation processes is exhausted by what is actually realized. The $\mathscr{I}$ values show that this process of exhaustion has proceeded to a greater extent for unmarked *-er* than for marked *-ster*, even though *-er* has the potential to show up with more types than *-ster*. As will be argued below, this reversal is the quantitative reflex of morphological markedness. In other words, even though the $\mathscr{I}$-based ranking is incompatible with the $\hat{S}$-based ranking, the two rankings are equally valid because they highlight different aspects of productivity.

The discussion thus far has focussed on the number of possible types S and the ways (absolute or relative to V) in which S can be used to evaluate morphological productivity. However, we may also approach productivity from a slightly different perspective, namely by focussing on the *probability* of encountering new types. In this approach, we distinguish between 'the types we have observed', i.e. the number of types V, and 'the probability that we will observe new types', say P. Obviously, both 'what we have' and 'what we may expect' should be taken into account when what Baayen and Lieber (1991) call the global productivity P^* of a word formation process is to be evaluated. To see this, consider two word formation processes E and F with identical probability P of giving rise to new types. Suppose that we have observed 50 types for E and 2000 types for F, then F is globally more productive than E. Similarly, if E and F show up with identical numbers of types, the process with the higher probability P will be the more productive of the two. In other words, an approach in which the notion of productivity is analyzed into the two components 'what we have' and 'the likelihood of additonal formations' requires a bi-dimensional analysis. Such an analysis is developed in Baayen (1989, 1991), where 'what we have' is quantified in terms of V relative to the overall corpus size, and where the probability P of encountering new types is expressed in terms of $\mathscr{P} = n_1/N$, the first derivative of the growth curve $V(N)$ of the relevant morphological category. The probability $\mathscr{P}$ can be understood as the quantitative formalization of the notion *degree* of productivity, following Bolinger's (1948: 18) characterization of this notion as "the statistical readiness with which an element enters into new combinations". Crucial to the proper understanding of $\mathscr{P}$ is that it measures only one component of global productivity, namely the probability of encountering new types. Its exclusive focus on the relevance of new types brings productivity rankings based on $\mathscr{P}$ only more or less in line with productivity rankings based on $\mathscr{I}$. Hence such rankings will differ from rankings of global productivity in terms of both $\mathscr{P}$ and V. In fact, analyses proceeding on $\mathscr{P}$ and V jointly will result in rankings that are similar to rankings based on $\hat{S}$ or $\hat{S} - V$.

1.2. *The empirical validity of* $\mathscr{P}$

We are now in the position to evaluate van Marle's claim that $\mathscr{P}$ gives rise to counterintuitive productivity rankings. In the light of the above, van Marle's claim should concern the validity of $\mathscr{P}$ as a measure of the probability with which new types may be expected *given that the number of types that have already been observed for the morphological category under consideration are not taken into account*. In other words, for van Marle's criticism to be valid, it should pertain to the theory-specific interpretation of $\mathscr{P}$ as a probabilistic formalization of that component of productivity that concerns the likelihood of encountering new types, where the notion 'new' is to be understood as new with respect to the types that have already been observed for a given category. Unfortunately, van Marle's objections against $\mathscr{P}$ do not bear on this issue at all. In fact, van Marle seems to have failed to grasp the fact that $\mathscr{P}$ focuses on only one component of global productivity. When arguing in a section on *global* productivity that the high $\mathscr{P}$-value for weakly productive *-erd* (0.444) is counterintuitive in the light of the $\mathscr{P}$-value of nominal compounds (0.225), van Marle fails to take into account the huge difference in the numbers of types, 6 for *-erd* and 4277 for the compounds. Conversely, in a section explicitly devoted to $\mathscr{P}$, van Marle suggests that the number of types V should have been taken into account. Unfortunately, V is irrelevant to the quantification of that component of global productivity that is captured by $\mathscr{P}$. Of course, V is highly relevant to the quantification of global productivity, which is the reason why a bi-dimensional analysis of global productivity was proposed in Baayen (1989, 1991). Van Marle's failure to distinguish between global productivity P^* and the theory-specific interpretation of the degree of productivity as a probabilistic notion expressed in terms of $\mathscr{P}$ is unfortunate. The issue that van Marle should have raised concerns the question whether speakers have intuitions that correlate with the numerical

values of $\mathscr{P}$ *when this statistic is considered without reference to* V. Now this appears to be the case only to the extent that when $\mathscr{P}$ becomes too low (0.013 for *-te*, 0.002 for *-eren* and 0.000 for English *en-*) an affix is felt to be unproductive. With intuitions concerning degrees of productivity being primarily linked to global productivity as meausred by $\mathscr{P}$ and V jointly, it remains a matter of debate to what extent the dissection of the notion of productivity into the two components V and $\mathscr{P}$ is psychologically valid. We will return to this issue in Section 2.

Interestingly, van Marle's attempts to understand why $\mathscr{P}$ fails as a supposed measure of global productivity lead him to raise a series of important issues that merit some discussion. One such issue concerns the representational validity of the corpus on the basis of which statistics such as V and $\mathscr{P}$ are calculated. Van Marle calls into question the reliability of the Eindhoven corpus (Uit den Boogaart 1975) that I have used for my productivity research for Dutch on the basis of its smallish size (600.000 tokens). If it is truly the case that the Eindhoven corpus is too small to be reliable, the results obtained on the basis of this corpus might suggest some quantitative trends at best. Fortunately, the reliability of the Eindhoven corpus need not be questioned with respect to the issue at hand. Even though the Eindhoven corpus is admittedly small,[3] the major patterns of productivity already emerge very clearly. It is only in the case of morphological categories that are represented by a very small number of tokens that special care is required for the interpretation of the statistics. As a case in point, consider *-erd*, for which we count 9 tokens representing 6 types, 4 of which are hapaxes. For such a small sample, the ratio n_1/N is an unreliable estimator of the rate at which new items may be expected to be encountered. For the sake of the argument we may ignore the mathematical niceties and simply proceed to compare the degree of productivity $\mathscr{P} = 0.444$ of *-erd* with that of the unproductive plural in *-eren*, for which $\mathscr{P}$ equals 0.002. Of the 15 nouns[4] which take this plural, 9 are found in the Eindhoven corpus. One of them occurs only once. Even a cursory inspection of the two frequency spectra shows that *-eren* is unproductive and that *-erd* may be productive, although some caution is required because of the extremely small number of tokens counted. Even though a larger corpus would have allowed a more precise evaluation of the productivity of *-erd*, the dominance of hapaxes and the absence of high-frequency types suggests at least some degree of productivity, in contrast to *-eren*. This evaluation is born out by the CELEX counts for the INL corpus: all 15 nouns which take the plurals in *-eren* are firmly attested with this plural, while 36 personal nouns in *-erd* are attested. Since the CELEX database has not registered the hapaxes, the observed number of 36 types is a lower bound only. Thus we find that the net increase in types is only 6 for *-eren*, but at the very least 30 for *-erd*. This illustrates that even for small samples $\mathscr{P}$ can be used to make qualitatively valid inferences about the magnitude of the number of unobserved types.

Another interesting issue raised by van Marle concerns the question how

$\mathscr{P}$ is arrived at. According to van Marle, the approach to morphological productivity developed in Baayen (1989, 1991) is motivated on the basis of the following observation, a quote from Baayen (1990a: 218):

> The larger the number n_1, the higher is the degree of productivity P. In general productive categories contain relatively few types with a high frequency. Consequently N remains relatively low, which also furthers a high value of $\mathscr{P}$. Conversely, categories with a low degree of productivity are characterized by a low number of hapaxes and a large number of types with a high frequency, and, consequently, by a low quotient n_1/N. (translation by van Marle 1991: 156)

Van Marle correctly observes that "it seems very much questionable whether it is justified to take the above observation — which seems to qualify as 'rule of thumb' at best — as a starting point for a formula with the general applicability of Baayen's $P = n_1/N$". He then proceeds to argue that the unproductive Dutch plural in *-eren* provides a counterexample in that the frequency spectrum of *-eren* ($V = 9$) is not characterized by many high-frequency types (only 1) while there are many low-frequency types (5). Unfortunately, van Marle misrepresents my position when he claims that the above observation is the starting point of my analysis. The starting point of my analysis is a probabilistic analysis of the frequency spectra of morphological categories (see e.g. Baayen 1991). The measure $\mathscr{P}$ is a non-parametric statistic that remains valid irrespective of the precise shape of the frequency spectrum. Retrospectively, the fact that the resulting quantitative measure is relevant to the issue at hand can be understood in terms of the abovementioned preponderance and scarcity of high and low frequency types respectively, notably so in the light of the psychological implications discussed in Baayen (1989, 1991), see also Section 2 below. In addition, Van Marle's claim that *-eren* has an atypical frequency spectrum for an unproductive suffix can only be maintained on the basis of his somewhat naive frequency grouping. A re-analysis using state-of-the-art criteria for frequency classes (Martin 1983) shows that 6 *-eren* plurals (66%) should be classified as frequent or very frequent, and that the remaining 3 plurals (only one of which is a hapax) should be classified as neutral, that is, as neither frequent nor infrequent. By way of comparison we may note that productive *-heid* shows up with only 65 formations (14%) that can be classified as very frequent or frequent, 401 types falling into the neutral frequency class.[5] This is not to say, of course, that atypical spectra cannot arise. In fact, such a case (English *re-*) is discussed at length in Baayen and Lieber (1991: 832—836). Even though the interpretation of $\mathscr{P}$ remains valid here, the usefulness of $\mathscr{P}$ as a productivity measure is stretched to its limits. In Section 2 it will be shown how this problem can be solved.

A methodological issue that is of interest here is the nature of statistical inference. In the quantitative approach advocated here, hapaxes are used to derive predictions about the likelihood of neologisms. An illustrative example of the misunderstandings that arise when this fact is not grasped is van Marle (1991: 157):

> ... note also that this non-productive category in *-eren* is represented by one hapax in the corpus, and that it has the potential to show up with 6 hapaxes in a new, larger corpus. As these hapaxes (would) have absolutely nothing to do with productivity, this also calls into question Baayen's general starting-point according to which hapaxes are always considered indicators of productivity. That many hapaxes do not have anything to do with morphological productivity is also directly confirmed by Uit den Boogaart (1975): only a quick glance reveals that many complex words that are attested only once in this corpus are perfectly common, actual words of Dutch.

Aside from the fact that it is extremely unlikely that a larger sample has the 'potential to show up with 6 hapaxes' — the number of hapaxes is a decreasing function of N, as described in detail in Baayen (1991) — van Marle equates hapaxes with neologisms. This is indeed counterproductive: especially in a small corpus, most if not all hapaxes are well-established items of the language. Once the notions hapax legomenon and neologism are carefully distinguished, however, the supposed incongruity between the item familiarity (Meijs 1985) of most hapaxes in the Eindhoven corpus and $\mathscr{P}$ as a measure of new types disappears. The number of hapaxes is used to estimate the likelihood of encountering types that have not been registered previously in the counting process. In the initial stages of this counting process, the hapaxes will generally be item-familiar. It is only for the larger corpora that neologisms will begin to appear, predominantly among the hapaxes. Even then, many hapaxes will be well-known lexical items. However, as the sample size increases, the proportion of the neologisms among the hapaxes will increase. Hence the probability of encountering neologisms is measured indirectly by means of the probability of encountering hapaxes. It is in this sense that $\mathscr{P}$ is a measure for the degree of productivity.[6]

Perhaps the most important issue raised by Van Marle (1991) concerns the role of language structure in the quantitative analysis of productivity. According to van Marle, a neglect of the purely systematic dimension of morphological structure is at least in part if not mainly responsible for the supposed problematic aspects of the present approach. He attempts to motivate this claim on the basis of the Dutch de-adjectival suffix *-te*. In Baayen (1989) a simple neural network was used to model the paradigmatics of the rival suffixes *-te* and *-heid*. The simulation results obtained strongly suggested that a morphological restriction on suffixation of *-te*, namely that it cannot attach to complex adjectives, should be re-analyzed as being the result of a phonological 'conspiracy'. Van Marle (1991), coming to the defence of the morphological restriction, argues that the fact that *-te* does not attach to complex adjectives should be accounted for by a competence restriction because newly constructed formations violating this condition are intuitively felt to be much worse than 'new' formations where *-te* is attached to a simplex base. Unfortunately, van Marle's argument is logically untenable. The problem is that the suffix *-te* is unproductive (Schultink 1962; Baayen 1989), a fact explicitly acknowledged by van Marle (1991: 160). Within a theory of lexical competence as envisaged by van Marle (1985, 1991), a statement to the extent that affix X is unproductive implies that no competence restrictions pertaining to rule X can be invoked to explain distributional patterns or intuitions concerning structural wellformedness. Hence the invocation of a morphological competence restriction to explain the impossibility of attaching *-te* to complex input is self-contradictory. Such a competence restriction is irrelevant in the same way as a competence restriction barring the unproductive plural suffix *-eren* from attaching to complex nouns would be unmotivated. Within the theoretical framework of van Marle (1985, 1991) formations in *-te* should be listed in exactly the same way as the unproductive plurals in *-eren* are listed. In other words, if *-te* is indeed unproductive, the removal of the competence restriction invoked by van Marle is a logical necessity. From this point of view, the fact that the network of Baayen (1989) categorially rules out the possibility of *-te* attaching to a suffixed adjective suggests that the strong intuitions concerning the unacceptability of such words may well arise on the basis of phonological constraints.[7] This is not a matter of neglecting the relevance of morphological structure for productivity, instead, it is an attempt to work out the consequences of a negative productivity verdict for the linguistic analysis of *-te*.

1.3. *Measuring global productivity: $\mathscr{P}^*$*

Having established that the global productivity P^* should be evaluated in terms of both $\mathscr{P}$ and V, we are nevertheless left with one serious problem: the functional relation g specifying how $P^* = g(\mathscr{P}, V)$ is to be computed from $\mathscr{P}$ and V is left unspecified. Although the coordinates of morphological categories in the $\mathscr{P} - V$ plane give some rough idea of how the two components of productivity are balanced, it is not possible to measure P^* exactly. For instance, the Euclidean distance to the origin in the $\mathscr{P} - V$ plane cannot be used in the absence of a principled way of scaling the horizontal and vertical axis. As a case in point, consider Dutch personal names in *-er* (0.076, 299) and *-ster* (0.231, 30) in the Eindhoven corpus. The unmarked case, *-er*, appears with the higher V but the lower $\mathscr{P}$. The reverse holds for the marked affix. Two questions are relevant here. First, which affix has the higher extent of global productivity? Should we give more weight to V, or is $\mathscr{P}$ more important? Second, isn't it rather counterintuitive to find that the marked category appears with the higher degree of productivity $\mathscr{P}$? Let's consider these questions in turn.

As to the relative weights of V and $\mathscr{P}$, recall that $\mathscr{P}$ is the first derivative of V as a function of N (see e.g. Baayen 1991). This implies that the weight of $\mathscr{P}$ should decrease with V: a large $\mathscr{P}$ in combination with a large V implies that more new types may be expected that when a large $\mathscr{P}$ co-occurs with a small V. In this sense, the high value of $\mathscr{P}$ for *-ster* is severely moderated by the small number of types. Even in the absence of $g(\mathscr{P}, V)$, the relative weight of $\mathscr{P}$ can, to some extent, be gauged.

Turning to the relation between productivity and markedness, we find that the higher degree of productivity for the marked suffix is precisely the quantitative reflex of markedness expected under the formalization developed

here. To see this, consider the analysis of the relation between markedness and frequency put forward by Martin (1988: 147):

> (Morphologically) marked items will be more restricted in meaning because they have (an) additional semantic feature(s) and cannot be used as generic items, consequently, all other things being equal, the unmarked item of a pair should belong to a higher frequency class.

Personal nouns in *-er* (median frequency 2 in the UdB corpus) will often be used in cases where the marked corresponding female personal noun in *-ster* (median frequency 1 in the UdB corpus) might have been used. This introduces a bias towards a preponderance of lower frequency types in the frequency spectrum of *-ster*, positively influencing its degree of productivity $\mathscr{P}$. What we find, then, is that $\mathscr{P}$ taps into the same aspect of productivity that is measured by $\mathscr{I}$, the ratio of possible to actual types S/V. Like $\mathscr{I}$, $\mathscr{P}$ ranks *-ster* above *-er*, even though $\hat{S}$ is very much larger for *-er* than for *-ster*. This reversal can be viewed as the quantitative correlate of morphological markedness. The unmarked category *-er* is found to have the larger number of potential types, but since it is also put to use to a far greater extent than its marked counterpart, its number of potential types exceeds the number of observed types to a lesser extent than is the case for *-ster*. Conversely, the marked status of *-ster*, which allows *-er* to pre-empt *-ster*, in combination with its productivity gives rise to the fact that the actual types realize a smaller proportion of the possible types. Hence the probability of sampling a type that is new with respect to the set of types that have already been observed is greater.

Nevertheless, the $\mathscr{P}$-based analysis has one serious drawback, namely that one would expect the 'general probability' of coming across new items in unmarked *-er* to be greater than the corresponding probability for marked *-ster*. The problem is that the probability $\mathscr{P}$ is based exclusively on the frequency spectrum of one particular morphological category only. Hence it expresses a probability that has the morphological category as its frame of reference. However, *-er* appears with 1676 tokens in a corpus with size N_C = 600,000 tokens while *-ster* appears with only 78 tokens. Consequently the average number of tokens of the corpus that has to be processed for a type with the required morphological constituency to appear is much larger for *-ster* (7692) than for *-er* (358). These differences in the so-called (mean) interarrival time $i = N_C/N$ are not taken into account by $\mathscr{P}$. They are indirectly taken into account in the evaluation of global productivity, which is assessed in terms of $\mathscr{P}$ and V, where $V = V(N)$ is a function of N. Unfortunately, the evaluation of global productivity in terms of $\mathscr{P}$ and V remains too impressionistic. Hence a probabilistic measure of global productivity that takes into account the interarrival time i would be of interest. More specifically, such a measure may be expected to give rise to productivity rankings similar to those obtained on the basis of S rather than $\mathscr{I}$, since the calculation of S also involves some form of standardization, V being considered in the limiting value of N ($S = \lim_{N \to \infty} V(N)$). To obtain such a measure we first consider the nature of $\mathscr{P}$ in some more detail.

When viewed from the perspective of the corpus as a whole, the probability $\mathcal{P}$ appears as a conditional probability, namely the probability that the next token sampled when the corpus is extended with one extra token represents a new type, given that this token belongs to the morphological category at hand. Because the probability of sampling new types is defined for each morphological category separately, differences in interarrival time are not taken into account. One way to standardize this interarrival time for all categories is to calculate the probability that a type belonging to some morphological category E is sampled, conditional on that this type has not been sampled before.[8] In other words, we now define the notion 'degree of productivity', understood as the likelihood of observing new types, as the conditional probability that the next token sampled belongs to the required morphological category, given that we know that this token represents a new type. Since the interarrival time for new types of arbitrary constituency is equal for all morphological categories, the desired standardization of interarrival times is obtained. Denoting the number of hapaxes observed for category E after t tokens of the corpus have been sampled by $n_{1,E,t}$, and denoting the total number of hapaxes of arbitrary constituency in these t observations by h_t, we find that the required conditional probability, say $\mathcal{P}^*$, equals $n_{1,E,t}/h_t$. In what follows I will refer to $\mathcal{P}$ as the category-conditioned degree of productivity and to $\mathcal{P}^*$ as the hapax-conditioned degree of productivity.

More formally, let t denote the size of the corpus as it is sampled thus far, that is, t is the number of tokens of the corpus that have already been encountered in the sampling process. We now sample one additional token, $X(t+1)$. Now let $\{X(t+1): f_{\mu_{X(t+1)},t} = 0\}$ denote the event that token $X(t+1)$ represents a type μ with frequency $f_\mu = 0$ at time t, and let $\{\mu_{X(t+1)} \in E\}$ denote the event that $\mu_{X(t+1)}$ belongs to morphological category E. Then $\mathcal{P}_{E,t+1}$, the probability that the token sampled at $t+1$ represents a new type μ, given that this token belongs to E, is found to equal $n_{1,E,t}/N_{E,t}$, the number of hapaxes belonging to E at time t divided by the total number of tokens of E sampled at time t:

$$(1)\qquad \mathcal{P}_{E,t+1} = \Pr(\{X(t+1): f_{\mu_{X(t+1)},t} = 0\} \mid \{\mu_{X(t+1)} \in E\})$$
$$= \frac{\Pr(\{X(t+1): f_{\mu_{X(t+1)},t} = 0\} \cap \{\mu_{X(t+1)} \in E\})}{\Pr(\{\mu_{X(t+1)} \in E\})}$$
$$\approx \frac{n_{1,E,t}}{t}\left(\frac{N_{E,t}}{t}\right)^{-1}$$
$$= \frac{n_{1,E,t}}{N_{E,t}}.$$

Note that the corpus size t is factored out in (1). Also note that differences in the mean interarrival time $i_E = t/N_{E,t}$ are not taken into account. For

instance, two categories E and F may appear with equal $\mathcal{P}$, nevertheless, if $N_{E,t} > N_{F,t}$ one will on average have to read through a larger part of the corpus in order to sample a type of category F than is the case for E, i.e. $t/N_{E,t} < t/N_{F,t}$.

In order to standardize the interarrival time for all categories E_i we condition on the event $\{X(t+1): f_{\mu_{X(t+1)},t} = 0\}$, the event that a new type is sampled, irrespective of the categorial membership of this type. With $h_t = \Sigma_i\, n_{1,E_i,t}$ the overall number of hapaxes this leads to

$$(2)\qquad \mathcal{P}^* = \Pr(\{\mu_{X(t+1)} \in E\} \mid \{f_{\mu_{X(t+1)},t} = 0\})$$
$$= \frac{\Pr(\{\mu_{X(t+1)} \in E\} \cap \{f_{\mu_{X(t+1)},t} = 0\})}{\Pr(\{f_{\mu_{X(t+1)},t} = 0\})}$$
$$\approx \frac{n_{1,E,t}}{t}\left(\frac{h_t}{t}\right)^{-1}$$
$$= \frac{n_{1,E,t}}{h_t}.$$

Note that $\mathcal{P}^*$ can also be viewed as measuring the relative contribution of a given morphological category to the overall vocabulary growth h_t/t.

Table 3 illustrates the kind of productivity rankings one obtains on the basis of $\mathcal{P}^*$. Generally, the productive categories are ranked in a satisfactory way. For instance, *-ster* is ranked below *-er*, as should be the case for a

Table 3. $\mathcal{P}^*$-based productivity rankings for noun-forming Dutch and English word formation processes. The English data are based on Baayen and Lieber (1991), the Dutch data on Baayen (1989, 1991). The numbers of hapaxes for Dutch and English are denoted by h_D and h_E respectively. The English productivity judgements are based on Baayen and Lieber (1991) and Marchand (1969: 245–247, 306–308).

Dutch			English		
category	$\mathcal{P}^* \cdot h_D$	prod.	category	$\mathcal{P}^* \cdot h_E$	prod.
nominal compounds	2591	+	simplex N	256	–
-tje	654	+	*-ness*	77	+
-ing	302	+	*-ation*	47	+
simplex N	294	–	*-er*	40	+
-heid	256	+	*-ity*	29	+
-er	128	+	*-ment*	9	±
-sel	21	+	*-ian*	4	±
-ster	18	+	*-ism*	4	+
-te	10	–	*-al*	3	±
-nis	6	–	*-ee*	2	±
-erd	4	+			

formalization for the global degree of productivity. Note, however, that the sets of simplex nouns appear with very high $\mathscr{P}^*$ values for both languages. Clearly the simplex categories cannot serve as a baseline condition for evaluating whether a category is productive, as in the case of $\mathscr{P}$. Also note that productive *-erd* is ranked below unproductive *-nis* and *-te*, that *-ster* is ranked below *-sel*, and that English *-ee*, which appears with the highest $\mathscr{P}$ in Table 3 of Baayen and Lieber (1991), is given the lowest ranking here. These reversals suggest that $\mathscr{P}$ and $\mathscr{P}^*$ are complementary measures, the primary use of $\mathscr{P}$ being to distinguish between unproductive and productive processes as such, $\mathscr{P}^*$ being especially suited to ranking productive affixes.

It is useful to consider the rationales for the two statistics developed here in some more detail. The rationale for the category-conditioned degree of productivity is rooted in the morphological category. In the same way as qualitative analyses of productivity have focussed on the question what base words can serve as input to a word formation process given that affix E is involved, the category-conditioned degree of productivity is developed in answer to the question what the probability of a new type is given that this new type belongs to the relevant category. This allows $\mathscr{P}$ to register the effects of transparency, which is known to be positively correlated with frequency. Categories with less transparent items will, due to this correlation, show up with more high frequency types, thereby lowering the category-conditioned degree of productivity. For instance, *-ster* feels more productive than *-sel*. This is not captured by $\mathscr{P}^*$, which assigns both affixes an approximately equal rank. Conversely, $\mathscr{P}$ homes in on this difference, emphasizing the transparency of *-ster* compared with the more subtle semantics of *-sel* (see Taeldeman 1990). However, since $\mathscr{P}$ is intimately linked with the morphological categories for which it is calculated, it allows cross-category comparisons only to a limited extent, requiring the complementation by V for the proper evaluation of the degree of global productivity.

What about the rationale for $\mathscr{P}^*$? Although this productivity measure appears to be a valuable tool, we should inquire what factors allow $\mathscr{P}^*$ to give rise to these rankings, the more so since $\mathscr{P}^*$ is linked to the morphological category only through the hapaxes, whereas $\mathscr{P}$ takes into account *all* types of the morphological category by means of their summed frequencies, N. Fortunately, $\mathscr{P}^*$ is supported by a processing rationale that builds on the morphological race model developed by Frauenfelder and Schreuder (1991).

2. FREQUENCY, TRANSPARENCY AND PRODUCTIVITY

In order to understand why $\mathscr{P}^*$ is able to provide satisfactory productivity rankings, we need a theory of lexical processing that addresses the question of how novel forms are understood. Such a theory is developed by Frauenfelder and Schreuder (1991). In Section 2.1 their morphological race model (MRM) is outlined. Section 2.2 discusses the consequences of this model for the statistics of productivity. It is shown that $\mathscr{P}^*$ emerges as the limiting form of a more general productivity measure $\mathscr{A}$ that estimates the activation levels of the access representations of affixes in the mental lexicon. Section 2.3 addresses the question to what extent productivity is determined by transparency. It is argued that Frauenfelder and Schreuder (1991) tend to overestimate the role of transparency. However, since the importance of transparency with respect to productivity cannot be denied, and since $\mathscr{P}^*$, unlike $\mathscr{P}$, does not take transparency into account, it is claimed that the productivity measure $\mathscr{A}$ is superior to $\mathscr{P}^*$. This claim is put to the test in Section 2.4.

2.1. *The morphological race model*

Frauenfelder and Schreuder (1991) develop a dual route model for lexical access. According to their morphological race model, words are processed by two routes, a direct route and a route involving morphological parsing. The direct route makes use of whole-word access representations, the parsing route employs access representations of stems and affixes, as in Caramazza *et al.* (1988). The two access strategies are assumed to run in parallel, the first route to complete lexical retrieval wins the race. Frauenfelder and Schreuder, like Stemberger and MacWhinney (1986, 1988), allow for some temporal overlap between the two routes, such that the highest frequency words will be effectively retrieved by the whole word address procedure, while for low-frequency words both routes have a chance of winning the race. Differences in degree of productivity are coded into the resting activation levels of the affix (and base word) access representations. Thus a high degree of productivity is modelled by a high resting activation level of the affix, speeding up the recognition of the affix and subsequent combination with the base word. This is approximately the position outlined in Baayen (1989, 1991).

Thus far the problem of how the relation between the frequency spectrum of morphological category E, the productivity of E and the activation level A_E of E should be made explicit is left unsolved. A simple mapping of N_E onto A_E is not a valid option. This would imply that the activation level of English *-ity* ($N = 42252$) should be higher than that of *-ness* ($N = 17481$), leading to the counterintuitive prediction that *-ity* have a real-time processing advantage with respect to *-ness*. Alternatively, one might consider writing $\mathscr{P}$ into the activation level. Apart from the problem that it is entirely unclear how such activation levels would develop over time and how they are arrived at, we run into the problem that the global degree of productivity cannot be evaluated in terms of $\mathscr{P}$ only.

Interestingly, the model developed by Frauenfelder and Schreuder (1991) allows this problem to receive a first approximative solution. Frauenfelder and Schreuder (1991) enrich the dual route model of lexical processing outlined above with the notion of the *early successful parse*. What they suggest is that the activation level A_E should be viewed as a function of the number of successful parses that complete before the whole-word address procedure. Given that the speed of the direct access route is determined by whole-word frequency, and assuming that the speed of the parsing route is determined by transparency, only the lower frequency regular words have a non-negligible probability of being effectively retrieved on the basis of the parsing route. This approach to lexical processing enables Frauenfelder and Schreuder (1991) to account for both the fact that storage appears to be pervasive in derivational morphology, at least for languages such as English and Dutch, and for the fact that novel forms can be processed adequately without having to appeal to some kind of back-up or fall-back procedure for morphological processing that comes into play only when lexical retrieval fails (Butterworth 1983).

2.2. *Degrees of productivity in the MRM*

Frauenfelder and Schreuder (1991) do not discuss the problem of how to account for the fact that word formation processes may have varying degrees of productivity. Nevertheless, their model allows such differences to be modelled in terms of differences in the activation levels of the access representations of affixes. As a first approximation, we may define a frequency threshold θ such that words μ for which $f_\mu \geqslant \theta$ are processed by the direct route while words with $f_\mu < \theta$ are parsed. The number of times that the parser is involved in the processing of complex words is given by the summed token frequencies of all types of category E for which the frequency is less than θ. Assuming that these low frequency words are all regular, this token count represents the number of positive re-enforcements for the activation level of affix E. Hence, this activation level $\mathscr{A}_E$ can be approximated in terms of this restricted frequency count. This leads to

(3) $$\mathscr{A}_{E,t} = \sum_{r=1}^{\theta-1} m_{r,E,t},$$

where $n_{r,E,t}$ denotes the number of types $\mu \in E$ with frequency r at time t and $m_{r,E,t}$ the number of tokens contributed by the $n_{r,E,t}$ types with frequency r to the total number of tokens. The probabilistic correlate of $\mathscr{A}_E$ is the relative frequency with which the relevant tokens appear in the corpus,

(4) $$\Pr(\mathscr{A}_{E,t}) = \frac{\sum_{r=1}^{\theta-1} m_{r,E,t}}{t}.$$

The ratio of these probabilities for categories E and F is given by

(5) $$\Pr(\mathscr{A}_{E,t}) : \Pr(\mathscr{A}_{F,t}) = \frac{\sum_{r=1}^{\theta-1} m_{r,E,t}}{\sum_{r=1}^{\theta-1} m_{r,F,t}}.$$

Interestingly, the ratio of the corresponding $\mathscr{P}^*$ values is

(6) $$\mathscr{P}^*_{E,t} : \mathscr{P}^*_{F,t} = \frac{n_{1,E,t}}{n_{1,F,t}}.$$

Combining (5) and (6) we find that $\Pr(\mathscr{A}_{E,t})$ can be considered as a (differently normalized) generalization of $\mathscr{P}^*_{E,t}$. $\Pr(\mathscr{A}_{E,t})$ is a processing estimate of the probability that the activation level of the affix is updated, $\mathscr{P}^*_{E,t}$ estimates the probability that the parser has to be invoked for the processing of neologisms. This leads to the rationale for $\mathscr{P}^*$ as the limiting value of $\Pr(\mathscr{A}_E)$ for $\theta \downarrow 1$. Note that for $\theta = 1$ we have the case in which hapaxes, which for $\theta = 1$ are accessed by means of the direct route, are used to estimate the probability of neologisms, for which the parsing route is the only option available. Observe, however, that although fixing θ at unity leads to a linguistically motivated formalization of the notion (global) degree of productivity in that it focuses on *new* as opposed to *old* types, it is psychologically less plausible in that the processing consequence of fixing θ at unity is that the activation level is supposedly raised only for the initial successful parse of a formation. This 'once only' view runs into the difficulty that V is not a reliable measure of productivity: some unproductive processes appear with more types than productive ones. Hence, for $\theta > 1$, $\Pr(\mathscr{A}_{E,\theta})$ or equivalently $\mathscr{A}_{E,\theta}$ is likely to be superior to $\mathscr{P}^*$. Before considering the resulting predictions we should first examine how the MRM handles the relations between transparency, pseudo-affixation and productivity.

2.3. *Transparency, pseudo-affixation and productivity*

Frauenfelder and Schreuder (1991) call attention to the effect of phonological (Cutler 1980, 1981) and semantic (Aronoff 1976) transparency on the processing of complex forms. They argue that phonological transparency — the extent to which the derived word is similar to its base in terms of e.g. vowel quality and stress pattern — influences the speed with which the constituent access representations of the incoming complex word become active. Hence phonologically less transparent formations are, other things being equal, predicted to require longer parsing times. With respect to semantic transparency they claim that a high degree of compositionality positively affects the processing time needed for the integration of stem and affix.

To what extent do the two kinds of transparency determine productivity? Frauenfelder and Schreuder appear to view phonological and semantic transparency as sufficient conditions for productivity given their claim (1991: 174) that "the inherent properties [i.e., phonological and semantic transparency, RHB] of words resulting from different word formation processes determine the productivity of these processes". However, it is well known that regularity is a necessary but not a sufficient condition for productivity.

For instance, the Dutch de-adjectival suffix *-te* is unproductive or at best marginally productive even though it is phonologically and semantically regular and transparent.[9]

To my mind, Frauenfelder and Schreuder tend to overestimate the supposed negative effect of phonological opacity on productivity. For instance, they describe the English suffix *-ity* as being unproductive (1991: 174), which they trace to its lack of phonological transparency. However, although *-ness* is more productive than *-ity*, the latter suffix cannot be claimed to be unproductive. In fact, it is *-ity* and not *-ness* which is the more productive of the two for the subdomain of adjectives in *-able* and *-ible* (Baayen and Lieber 1991, see also Anshen and Aronoff 1988). Fortunately, Frauenfelder and Schreuder considerably weaken their claim that phonological transparency is a strong factor determining productivity in their discussion of word recognition in Finnish (1991: 180—181). In this language, a high degree of morphological complexity goes hand in hand with a high degree of phonological opacity for many surface complex forms. This, however, does not seem to slow down comprehension, suggesting that the speed of the parser is not seriously affected by phonological opacity as such — regular phonology need not be detrimental to morphological processing.

Next consider semantic transparency and productivity. Frauenfelder and Schreuder (1991: 176) argue that the "semantic coherence of a form affects the time taken to integrate the meanings of its stem and affixes." Unfortunately, this way of modelling the negative effect of a lack of compositionality on word recognition is linguistically implausible. Consider the processing of the Dutch verb *her-haal*, 'to repeat', in which the productive and fully transparent prefix *her-*, compareable with English *re-* in *restate*, combines with the verb stem *haal*, 'to fetch'. At any time the compositional reading 'to fetch again' is available in principle, even though the non-compositional reading 'to repeat' will be more often retrieved. A similar situation obtains for English *return*. It is not clear in what way the non-compositionality of the whole word *herhaal* should affect the parsing of *her-* followed by *haal*. Both constituents have their own meaning, the integration of which into the reading 'to fetch again' should proceed independent of the reading 'to repeat' of the stored complex form *herhaal*. What this implies is that the negative effect of non-compositionality on the resting activation level of the affix should not be modelled into the parser itself — by definition the parser should produce only compositional readings — but rather into a post-retrieval process which evaluates the felicity of the semantic reading delivered by the winning route.

In the case at hand, we have two possible classes of context, one requiring the reading 'to fetch again', the other the reading 'to repeat'. No problems arise in the case that *herhaal* is a high frequency verb occuring in a context which requires the non-compositional reading. The same holds for the opposite case, the verb having a low frequency in a compositional context. The remaining two logical possibilities require further discussion. If the context requires the compositional reading for a high-frequency verb of the type *herhaal*, the output of the whole word route delivering the non-compositional reading has to be rejected. In this case the system has to wait for the compositional reading to be delivered by the parsing route. Given that the parser is able to deliver the required compositional reading in time, the activation levels of *her-* and *haal* will be increased. If, however, the context requires the non-compositional reading while the verb has a low frequency, the output of the whole-word route has to be waited for in combination with negative re-enforcement of the constituent resting activation levels. In other words, a lack of transparency negatively affects the speed of the parsing route, not because of intrinsic computational difficulties in obtaining the non-compositional reading, but indirectly by negative re-enforcement flowing back into the lexicon by post-retrieval processes.

A similar situation may well obtain in the case of pseudo-affixation. Some, albeit not unequivocal experimental evidence for extra processing costs for strings like *reveal*, where both *re-* and *veal* have access representations but where there is no compositionality whatsoever, has been advanced experimentally by Taft and Forster (1975) and Taft (1988) for English. Recently, Laudanna and Burani[10] obtained experimental results suggesting that in Italian the processing of otherwise equally productive affixes in legal non-words is affected by the number of pseudo-affixed strings in the language, the processing of the affix with the larger number of pseudo-affixed words being slowed down significantly.

How can such findings be incorporated within the morphological race model? As a first approximation, we may assume that only those pseudo-affixed words that have a frequency less than the frequency threshold θ should be taken into consideration. It is only for these words that the parser is likely to deliver some interpretation, for instance 'to repeat an action involving veal" in the case of *reveal*, before the whole word address route completes its retrieval of the correct reading. As in the case of non-compositional formations discussed above, the infelicity of the computed meaning in context necessitates waiting for the output of the whole-word route in combination with negative feedback with respect to the parsing route.

Finally note that semantic transparency, like phonological transparency, is a necessary but not a sufficient condition for productivity. The Dutch plural in *-eren* is unproductive, nevertheless it is semantically fully transparent. When we compare productive affixes with respect to their degree of productivity, we again observe differences that cannot be traced to phonological or semantic transparency. For instance, female personal nouns in *-ster* are fully transparent as to form and meaning. The fact that *-ster* is less productive than its unmarked counterpart *-er* or de-adjectival *-heid* cannot be traced to the processing factors discussed by Frauenfelder and Schreuder (1991). Differences in the usefulness of items in *-ster*, *-er* and *-heid* to the language community, differences in markedness, the effects of paradigmatic rivalry, but also social convention as such — Dutch *-ster* is much less productive than its German counterpart *-in* — should not be neglected. These factors are reflected in the frequencies of use. In fact, it is only type and token frequency, not regularity in the sense of phonological and semantic transparency, that distinguishes between unproductive *-eren* and productive *-ster* and between unproductive *-te* and productive *-heid*. As we shall see below, this frequential difference has radical consequences for lexical processing. Facts such as these suggests that token frequencies by themselves, reflecting the amount of exposure to a given morphological pattern, co-determine the speed of the parsing route. As we shall see below, this is exactly what is predicted on the basis of the MRM.

2.4. *Predictions*

The estimates of $\mathscr{P}$ and $\mathscr{P}^*$ discussed in the previous sections are based on the notion of the morphological category. That is, only those complex words are taken into account in which the affixed word is not itself embedded in some larger word. For instance, when the category of denominal verbs with *re-* is considered, *reforestation* is not taken into account along with *reforest*. It is of course also possible to base one's calculations on the cumulated frequencies of all words in which a given member of a given morphological category occurs. In that case the token frequency of *reforest* is obtained by summation over the frequencies of *reforest, reforestation*, ... It is this way of counting tokens that I will use for estimating $\mathscr{A}$.

There are two reasons for not using the cumulated token frequencies when calculating $\mathscr{P}$ or $\mathscr{P}^*$. First, consider a compound like *belastingbetaler*, 'tax payer', where *belasting*, 'tax', is derived from the verb *belasten* by suffixation of *-ing*, and where *betaler* is the agent noun in *-er* of the verb *betalen*, 'to pay'. By itself, the noun *betaler* occurs 10 times in the INL corpus, whereas the noun *belastingbetaler* occurs 81 times. The frequency of *belasting* equals 1569. Although one might collaps the frequencies of *betaler* and *belastingbetaler*, it is unclear to what extent the frequency of *belastingbetaler* is shaped by synthetic compounding and suffixation with *-ing* rather than by suffixation with *-er*. In the light of the difficulty of teasing apart the frequential weights of the various word formation processes involved in multiply complex words, I have opted to base $\mathscr{P}$ and $\mathscr{P}^*$ on the morphological category. In the case of $\mathscr{P}^*$ this has the additional advantage that the values of $\mathscr{P}^*$ can be interpreted directly as the relative contributions of the associated word formation processes to the growth curve of the vocabulary as a whole. When a morphologically complex hapax is counted separately for each of the morphological categories to which its constituents or itself may belong, this is no longer possible.

Second, if one's focus is to be on the word as independent free forms of the lexicon, then it would be counterproductive to cumulate token frequencies for formations such as *voorzienigheid*, 'providence' and *linksigheid*, 'leftishness'. In the case of *voorzienigheid* (INL frequency 126) the base word *voorzienig* is extremely rare (INL frequency 3). By bringing together the frequencies of both *voorzienig* and *voorzienigheid* we would predict that *voorzienig* should intuitively feel to belong to the same frequency class as *voorzienigheid*, contrary to fact. The status of *voorzienig* as a lexical item of Dutch is very much weaker than that of *voorzienigheid*. Turning to *linksigheid*, itself a low frequency word (INL frequency 3), we find that its base word is not attested in the INL corpus at all. Although *linksig* is a possible independent word of Dutch, it is not actualized as such in the INL corpus. Hence it should not be counted as if it were an independently occuring item, even though it is embedded in the abstract noun *linksigheid*.

This is not to say that the existence of *linksigheid* is not indicative of the productivity of the adjective forming suffix *-ig*. In fact, when we shift our perspective from the lexical status of complex words to the way in which they are processed, it is more natural to assume that token frequencies should be accumulated on the condition that these frequencies do not exceed the frequency threshold θ. If, for example, θ is fixed at 40, *voorzienigheid* (126) is not parsed, nor is *belastingbetaler* (81). On the other hand, *betaler* (10), *voorzienig* (3) and *linksigheid* (3) are all parsed, with *linksigheid* contributing to the activation level of the suffix *-ig* at the stage where *linksig*, 'leftish', is parsed.

We are now in the position to calculate activation level based productivity rankings for some fixed threshold θ. Table 4 lists the results obtained for a selection of English affixes ($\theta = 8$), Table 5 lists the corresponding results for a series of Dutch affixes ($\theta = 20$). The threshold values for English and Dutch are chosen differently in order to take into account the differences in the size of the corpora which underlie the estimates presented here. The second column of these tabels (REG) lists the number of times a token is processed for which $f_\mu < \theta$. The third column (PSE) lists the number of pseudo-affixed tokens encountered for which $f_\mu < \theta$. Only those pseudo-affixed words are taken into account in which a pseudo-affix is adjacent to a lexical item. Thus *wristlet* is not counted as a (phonologically) pseudo-affixed word since there is no stem *stlet* in English. But *defilement, ingeneously, recap* and *retail* are genuine pseudo-affixed words that on the surface are similar to regular complex words such as *decipherable, inaudible* and *to resole*.[11] The fourth column (REG-PSE) lists the resulting estimates of the access representation resting activation levels. Negative values are set to zero, suggesting the loss of the corresponding affix representations. The last column lists the average number of tokens in the corpus that have to be processed for the activation level to be raised once. As in the case of $\mathscr{P}^*$, satisfactory productivity rankings are obtained. Note that the unproductive English prefixes *en-* and *be-* as well as the Dutch unproductive suffixes *-te* and *-eren* show up with extremely low activation levels that may even drop to zero for *en-* and *-eren* when the effect of pseudo-affixation is taken into

Table 4. Data on the occurrences of regular (REG) and pseudo-prefixed (PSE) wordforms in the CELEX-COBUILD corpus, phonological representation. $C = 17979343$. Frequency threshold: $\theta = 8$.

affix	REG	PSE	REG-PSE	C/(REG-PSE)
en-	9	30	0 (−21)	∞
be-	26	6	20	898967
re-	56	6	50	359587
de-	59	5	54	332951
mis-	111	0	111	161976
-al	170	27	143	125730
-ment	154	0	154	116749
-ish	156	0	156	115252
in-	344	151	193	93157
un-	234	0	234	76835
-able	314	0	314	57259
-ity	337	0	337	53351
-ness	791	0	791	22730

Table 5. Data on the occurrences of regular (REG) and pseudo-prefixed (PSE) wordforms in the CELEX-INL corpus, phonological representation. $C = 42380000$. Frequency threshold: $\theta = 20$.

affix	REG	PSE	REG-PSE	C/(REG-PSE)
-eren	3	159	0 (−156)	∞
-te	135	134	1	42380000
her-	207	0	207	204734
ont-	341	57	284	149225
-sel	379	19	360	117722
be-	747	147	600	70633
-ster	875	5	870	48712
-baar	1013	18	995	42592
ver-	2350	67	2283	18563
on-	3148	41	3107	13640
-heid	7357	8	7349	5766
-ing	8924	247	8677	4884

account, suggesting a complete dependency on the storage-based access route. Also note that even though *-ster* and *-eren* are equally transparent, only *-ster* appears with a reasonable activation level.

Of course, it should be kept in mind that the above calculations are rough approximations only. For instance, it is unclear whether the negative re-enforcements *PSE* should be given the same weight as the positive re-enforcements *REG*. Since it is likely that the computation of the semantic representation of the stem + affix combination is more time-consuming for anomalous strings of the type *reveal* than for regular complex words, the negative effect on the resting activation level is probably overestimated. Another problem concerns the appropriate choice of θ. There are no criteria available at present for guiding this choice. For the lower frequency ranges the productivity rankings obtained by varying θ appear to be highly similar. However, it may well turn out that θ need not be as low as is assumed here. The low choice of θ in the present paper is motivated by the constraint that only semantically transparent complex words contribute to the activation level A. Since transparency is inversely correlated with frequency, higher values of θ would lead to the inclusion of opaque and less transparent forms in the frequency counts. In the absence of indications in the CELEX database of the (degree of) semantic transparency of complex words, and in the absence of principled methods by means of which degrees of transparency and their effect on processing can be properly evaluated, the research strategy adopted here is to concentrate on that frequency range where complex words are most likely to be transparent.

Let's finally compare the different predictions made by $\mathscr{P}$, $\mathscr{P}^*$ and $\mathscr{A}$ for the prefixes *re-*, *de-* and *en-* and for the rival affixes *in-* and *un-*. First consider unproductive *en-*. Baayen and Lieber (1991: 822—823) pointed out that *en-* does not appear with even a single hapax in the Cobuild corpus, indicating that no new types with *en-* are to be expected — both $\mathscr{P}$ and $\mathscr{P}^*$ are zero. As shown in Table 4, *en-* emerges with the lowest number of regular parsings REG in the present study. In addition, there is a substantial number of tokens that contain an initial phonological sequence indistinguishable from *en-* where no live suffix of the language is involved, such as *inversion* and *infraction*. These pseudo-prefixed words may cause the activation level of the access representation of *en-* to become zero, confirming the earlier analysis in terms of $\mathscr{P}$ and V.

Next consider the prefixes *de-* and *re-*, both of which intuitively feel productive. However, in spite of its productivity, the latter prefix appears with an extremely low $\mathscr{P}$ value. This finding is discussed in detail in Baayen and Lieber (1991: 832—836). They argue that this low score is due not so much to the absence of low frequency types as to the presence of a number of atypical high-frequency types, the semantics of which are opaque. Nevertheless, they admit that the limits of the $\mathscr{P}$-based approach are reached. This problem is resolved in the present analysis, where *re-* and *de-* appear with roughly equal activation levels that are substantially higher than those of *be-* (marginally productive at best) and *en-* (unproductive). We may trace this reversal to two different factors. First, the high-frequency opaque formations in *re-* that contribute to the low value of $\mathscr{P}$ do not enter into the calculation of $\mathscr{A}$, allowing the low-frequency formations present in the distribution of *re-* to determine its degree of productivity. This is already reflected in the $\mathscr{P}^*$ values. While the $\mathscr{P}$ values for *de-* and *re-* reported in Baayen and Lieber (1991: 831) are 0.001590 and 0.000042 respectively, the $\mathscr{P}^*$ values are much more similar ($3/h$ and $1/h$ respectively). Second, the $\mathscr{A}$ value of *re-* is also positively affected by the fact that low-frequency complex words in which formations with *re-* are embedded are included in the frequency counts, bringing the estimated activation level of *re-* (50) to the same order of magnitude as that of *de-* (54). All this suggests that $\mathscr{A}$ and $\mathscr{P}^*$ are more sensitive to differences in productivity than $\mathscr{P}$.[12]

Next consider the rival prefixes *un-* and *in-*. In terms of $\mathscr{P}$, *un-* is slightly more productive than *in-* (0.0005 versus 0.0004), but in terms of V it is *in-* that appears with the higher score (237 versus 184 for *un-*). Interestingly, *un-* appears as the more productive prefix in the ranking of Table 4. This, however, is not so much due to the numbers of low-frequency regular occurrences — we count more such tokens for *in-* (344) than for *un-* (234). The reversal is brought about by the large numbers of words with initial /In/, such as e.g. *endurable* and *infusion*, where negative *in-* is not present. Since intuitively *un-* feels more productive than *in-*, even though *in-* appears with more types and more low-frequency tokens than *un-*, we tentatively conclude that an $\mathscr{A}$-based analysis in which the effects of pseudo-prefixation are taken into account is superior to one in which these effects are neglected.

Summing up, of the three productivity statistics discussed here it is $\mathscr{A}$ which appears to be the most reliable. Interestingly, $\mathscr{A}$ is the only productivity statistic that is psychologically motivated. The kind of knowledge tapped into when speakers make intuitive productivity judgements appears to be closely linked to the resting activation levels of affix access representations. This might explain why such intuitions are of an ordinal rather than of an interval nature, that is, why we have intuitions about whether affix a_1 is more productive than affix a_2 — given that there is a substantial difference in productivity — while we do not have intuitions concerning the exact number of actual or probable types with which a_1 exceeds a_2. Given the present theory, our ability to make these ordinal judgements is to be traced to the positive correlation between high activation levels and high numbers of actual (or possible) types. For instance, an affix like *-erd* that appears with 894 tokens in all is unlikely to have more types than a suffix like *-baar* (cf. English *-able*) for which the number of tokens below threshold level only already adds up to more than 1000. Given the Zipfian shape of word-frequency distributions, *-baar* is much more likely to have many types than *-erd*, as is indeed the case. Interval measurements are impossible, however, since activation levels are not in a direct functional relationship with V or S.

At the same time the present theory also explains why an affix may intuitively feel to be productive, even though the number of types attested in corpora or dictionaries may be quite small. The Dutch suffixes *her-* (English *re-*) and *-erd* illustrate this point: even though the numbers of types are relatively low (24 and 6 respectively in the UdB), the low token frequencies of these types require parsing, thereby protecting the resting activation levels of the affixes against decay. Finally note that the present analysis also explains why both V and $\mathscr{P}$ *by themselves* do not provide intuitively satisfying productivity rankings: the access representation resting activation levels on which intuitive judgements are based do not allow the analytical distinction between 'what we have' (V) and 'what we may expect' ($\mathscr{P}$) to be made.

The measures $\mathscr{P}^*$ and $\mathscr{A}$ have one other advantage above $\mathscr{P}$ that is worth mentioning, namely that they are more likely to be useful for the large corpora (t > 100,000,000) that are becoming available at present. The problem with $\mathscr{P}$ is that in the limit of $N \rightarrow \infty$ the ratio n_1/N will tend to zero. Hence the sensitivity of $\mathscr{P}$ to differences in productivity will decrease for very large N. This problem is avoided when $\mathscr{P}^*$ or A are used. In the case of $\mathscr{P}^*$, we may consider the inequality

$$(7) \qquad \lim_{t \rightarrow \infty} \frac{n_{1,e,t}}{h_t} > 0$$

to define fully productive word formation. In the case of $\mathscr{A}$, the threshold θ can be parameterized for t when comparisons across corpora or even languages are to be made, such that

$$(8) \qquad \frac{\theta_t}{t} = c,$$

with c some constant.

3. CONCLUSIONS

The present paper has addressed the question in what way productivity and frequency are related. In contrast to van Marle (1991), I have argued that productivity can be measured on the basis of data pertaining to language use, namely word frequencies. Depending on what aspect of productivity is of primary interest, we may either focus on the morphological category itself, and estimate the number of potential types S and the category-conditioned degree of productivity $\mathscr{P}$, or we may focus on the psychological status of word formation processes, in which case the hapax-conditioned degree of productivity $\mathscr{P}^*$ and the estimated activation level $\mathscr{A}$ are appropriate. The primary use of $\mathscr{P}$ is to distinguish between productive and unproductive word formation processes as such, whereas $\mathscr{P}^*$ and $\mathscr{A}$ are particularly suited to ranking productive processes according to their degree of productivity. In the light of its sensitivity to semantic transparency, the productivity statistic $\mathscr{A}$ emerges as an especially promising productivity measure.

NOTES

* I am indebted to Geert Booij, Mirjam Ernestus, Uli Frauenfelder, Matthias Hüning, Jan-Peter de Ruiter, Rob Schreuder, Ariane van Santen and an anonymous referee for many helpful comments and insights.

[1] These estimates of S differ from the estimates presented in Baayen (1989, 1991). The present estimates are obtained on the basis of the best significant fit ($p < 0.05$) of the available theoretical models (extended generalized Zipf's law (Khmaladze and Chitashvili 1989), Carroll's (1967) lognormal law and Sichel's (1986) generalized Gauss-Poisson law) to the data. Within the limits of an approach based on the (questionable) assumption that words occur independently in texts, the estimates given here are optimal. See Baayen (1992) for detailed discussion.

[2] Note that the problems residing in a ranking based on V, namely that *-ster* would be ranked below *-sel* even though *-sel* feels much less productive than *-ster*, are avoided.

[3] The reason that I have not made use of the 42,000,000 INL corpus resides in the fact that CELEX and the INL have failed to process the hapaxes in the INL corpus for their lexical database on the unfortunate a priori assumption that only words with either a minimal frequency of 2 or words with a minimal dispersion of 2 are of interest. In addition, the Eindhoven corpus has the advantage that it consists of a large number of very small text fragments (75 up to 300 words) rather than of full texts. Consequently the effects of topic continuity on the re-use of words within texts are considerably reduced, bringing the statistical analysis of morphologically defined subsets of words as they appear in the text fragments more (but not completely, see Baayen 1989) in line with the assumption of statistical independence underlying the present approach.

[4] The plural *klederen*, 'clothes', should be added to van Marle's (1991: 157) list.

[5] The difference between the two proportions (6/9, 65/466) is highly significant, $p < 0.001$.

[6] See Baayen and Lieber (1991: 812–814) for an in depth discussion of this issue.

[7] With respect to van Marle's (1991) objections against the supposedly wrong predictions produced by the simple pattern matcher discussed in Baayen (1989) it should be noted that all errors produced in simulation 1 are possible words in *-te* if *-te* is taken to be productive. The errors arise simply because the pattern matcher received limited trained on a highly restricted data set. The Eindhoven corpus is indeed too small to allow the model to master the arbitrary list of established formations in *-te*. In other words, the kind of errors made by the model are precisely the kind of errors made by children who have not yet completely mastered the language norm in the sense of Coseriu (1975), even though they have mastered the language system.

[8] Another way of standardization is discussed in Baayen (1989: 117–121), where I consider the effects on $\mathscr{P}$ of reducing the larger samples to the size of the smaller ones. This amounts to comparing affixes for equal N. The problem with this procedure, however, is that the additional types and tokens the larger samples show up with in one and the same overall corpus are ignored. The wish not to ignore the relevance of these extra types led me to develop the present bi-dimensional analysis of productivity, in which both differences in V and differences in the growth rates $\mathscr{P}$ are taken into account.

[9] No stress shift is involved, and only one item, *lengte*, derived from *lang*, 'long', shows up with a vocalic alternation. Similarly, only a single item is semantically opaque, namely *groente*, 'vegetable'.

[10] Paper read at a workshop on lexical statistics, MPI, November 1991.

[11] The present counts of low-frequency pseudo-affixed words assume that conditions on the word category of the base need not be met for a low-frequency pseudo-affixed word to effect negative feedback to the access representations of the relevant stems and affixes. Thus the Dutch monomorphemic word *route*, 'route', can be phonologically parsed as *roet-te*, 'sootness' and is counted as a pseudo-affix. This procedure is motivated on the ground that affix-generalization is not uncommon for the more productive affixes. For instance, although Dutch *-te* and *-heid* normally do not attach to nouns, one example from the Uit den Boogaart (1977) corpus is known to me where this is nevertheless the case, namely *goud-heid*, 'goldness'.

[12] In the case of *re-* and *de-*, only those formations have been taken into account in which the vowel of the prefix is realized as an /i/ rather than as /I/ or /ɛ/. The same holds for the corresponding pseudo-prefixed words. Interestingly, the fact that only semantically transparent prefixed words show up with /i/ — contrast *reline* with *reformation* — appears to be functionally motivated, for when we count all low frequency occurrences of formations in *re-* and *de-* regardless of the vowel quality of the prefix and substract the corresponding numbers of pseudo-prefixed words, the estimated activation levels become zero. In other words, the vowel that appears in the regular formations minimizes the negative effect on productivity of possible pseudo-prefixed words.

REFERENCES

Anshen, F. and M. Aronoff. 1988. "Producing Morphologically Complex Words". *Linguistics* 26, 641–655.

Aronoff, M. 1976. *Word Formation in Generative Grammar.* Cambridge, Mass.: MIT Press.

Baayen, R. H. 1989. *A Corpus-Based Approach to Morphological Productivity. Statistical Analysis and Psycholinguistic Interpretation.* Dissertation, Vrije Universiteit, Amsterdam.

Baayen, R. H. 1990a. "Corpusgebaseerd Onderzoek naar Morfologische Productiviteit". *Spektator* 19, 213–233.

Baayen, R. H. 1990b. De Graad van Produktiviteit van het Suffix -ing". *Forum der Letteren* 31, 203–217.

Baayen, R. H. 1991. "Quantitative Aspects of Morphological Productivity". In G. E. Booij and J. van Marle (eds.), *Yearbook of Morphology 1991.* Dordrecht: Kluwer Academic Publishers, 109–149.

Baayen, R. H. 1992. "Statistical Models for Word Frequency Distributions: A Linguistic Evaluation". To appear in *Computers and the Humanities.*

Baayen, R. H. and R. Lieber. 1991. "Productivity and English Derivation: A Corpus Based Study". *Linguistics* 29, 801–843.

Bolinger, D. L. 1948. "On Defining the Morpheme". In D. L. Bolinger, *Forms of English. Accent, Morpheme, Order.* Cambridge, Mass.: Harvard University Press, 183–189.

Butterworth, B. 1983. "Lexical Representation". In B. Butterworth (ed.), *Language Production (Vol. II): Development, Writing and Other Language Processes.* London: Academic Press, 257–294.

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 (eds.), *Computational Analysis of Present-Day American English.* Providence, Brown University Press, 406–424.

Coseriu, E. 1975. "System, Norm und Rede". In E. Coseriu, *Sprachtheorie und allgemeine Sprachwissenschaft.* München: Fink, 11–101.

Cutler, A. 1980. "Productivity in Word Formation". *Papers from the Sixteenth Regional Meeting of the Chicago Linguistic Society*, 45–51.

Cutler, A. 1981. "Degrees of Transparency in Word Formation". *CLS* 26, 73–77.

Frauenfelder, U. H. and R. Schreuder. 1991. "Constraining Psycholinguistic Models of Morphological Processing and Representation: The Role of Productivity". In G. E. Booij and J. van Marle (eds.), *Yearbook of Morphology 1991.* Dordrecht: Kluwer Academic Publishers, 165–183.

Khmaladze, E. V. and R. J. Chitashvili. 1989. "Statistical Analysis of Large Number of Rare Events and Related Problems". *Transactions of the Tbilisi Mathematical Institute* 91, 196–245.

Marchand, H. 1969. *The Categories and Types of Present-Day English Word Formation. A Synchronic-Diachronic Approach.* 2nd edn. München, C. H. Beck'sche Verlagsbuchhandlung.

Marle, J. van. 1985. *On the Pardigmatic Dimensions of Morphological Creativity.* Dordrecht: Foris.

Marle, J. van. 1991. "The Relationship Between Morphological Productivity and Frequency: A Comment on Baayen's Performance-Oriented Conception of Morphological Productivity".

In G. E. Booij and J. van Marle (eds.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer Academic Publishers, 151—163.

Martin, W. 1983. "On the Construction of a Basic Vocabulary". In S. Burton and D. Short (eds.), *Proceedings of the 6th International Conference on Computers and the Humanities*. Comp. Science Press, 1983, 410—414.

Martin, W. 1988. "Lexical Frequency". In K. van Reenen-Stein (ed.), *Distributions spatiales et temporelles, constellations des manuscrits: études de variation linguistique offertes à Anthonij Dees à l'occasion de son 60me anniversaire*. Amsterdam: Benjamins, 139—152.

Meijs, W. 1985. "Morfologische Regels en het Mentale Lexicon". *Glot* 8, 183—206.

Santen, A. van. 1992. *Produktiviteit in Taal en Taalgebruik*. Dissertation, Leiden University.

Schultink, H. 1962. *De Morfologische Valentie van het Ongelede Adjectief in Modern Nederlands*. The Hague: van Goor Zonen.

Sichel, H. A. 1986. "Word Frequency Distributions and Type-Token Characteristics". *Mathematical Scientist* 11, 45—72.

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.

Taeldeman, J. 1990. "Afleidingen op *-sel*: Semantiek, Produktiviteit en Integratie in een Globale Verantwoording van Deverbatieve Nomina". *Studia Germanica Gandensia* 19, 77—115.

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.

Uit den Boogaart, P. C. (ed.) 1975. *Woordfrequenties in Gesproken en Geschreven Nederlands*. Utrecht: Oosthoek, Scheltema and Holkema.

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

Review article

Morphology without word-internal constituents: a review of Stephen R. Anderson's *A-Morphous Morphology**

ANDREW CARSTAIRS-McCARTHY

1. INTRODUCTION

Stephen R. Anderson's first published contribution to morphological theo dates from 1977, and several subsequent articles in *Linguistic Inqui Natural Language and Linguistic Theory* and elsewhere have kept him in t morphological limelight ever since (Anderson 1977; 1982; 1984; 198 1988). No survey of current views of inflection can be complete witho mentioning his approach to morphosyntactic representations and his analy of verb forms in various languages, particularly Georgian. But only now h he published a full-length book on morphology (Anderson 1992). For reade of this yearbook, that in itself is important news. What makes the ne appear exciting as well as important is that Anderson's main avowed aim to challenge the morphological mainstream (both structuralist and gener tive) by propounding a morphology without morphemes, whether inflection or derivational.[1] Since at the same time Anderson maintains, quite conve tionally, that 'the object of study in morphology is the structure of words' (7), this challenge may strike some readers at once as paradoxical or ev contradictory. To slice a word or a sentence into morphemes is not alwa straightforward, admittedly; nevertheless, describing the structure of a co plex word (many would say) is inseparable from analysing it into its co stituent morphemes. How does Anderson seek to persuade us otherwise?

I share much of Anderson's outlook: sympathy for the traditional distin tion between inflection and derivation, suspicion of premature theorising the basis of a morphologically rather limited language such as English, and belief that at least some aspects of word structure are explicable only reference to factors quite distinct from any which affect either syntax phonology. Yet Anderson's answer to the question just posed seems to r curiously unpersuasive. It is almost as if, having hit on 'a-morphous mo phology' as a suitably memorable title for his approach (p. 1), Anderson th feels obliged to justify it. I will offer reasons for this criticism in Section But Anderson certainly raises some important questions about the extent which word-internal structure is visible, quite independently of the fram work in which he chooses to discuss these questions. I return to the questions in Section 3.

This article concentrates on what is most central to Anderson's 'a-mo phous' manifesto. This means that I will have nothing to say about mo