43. Corpus linguistics in morphology: morphological productivity

R. Harald Baayen

(Baayen, 43)

1. Morphological productivity

2. Theoretical frameworks

3. Measuring productivity

4. Forces shaping productivity

5. Concluding remarks

6. Literature (a selection)

1. Morphological productivity

The vocabulary of English and most other languages contains many words that have internal structure (see also Articles 26 and 31). Words such as STRANGENESS, WEAKNESS, and SOFTNESS contain the formative NESS, which is usually found to the right of adjectives. Words ending in NESS almost always are abstract nouns. We refer to the sets of words sharing aspects of form structure and aspects of meaning as morphological categories.

Some morphological categories have a fixed or declining membership, while others have a growing membership. Categories with fixed or declining membership are said to be unproductive, categories with growing membership are described as productive.

Morphological categories differ tremendously in size. Some contain only a few

words (e.g., the category of words in TH such as WARMTH and STRENGTH), others may have tens of thousands of members (e.g., nominal compounding). A large morphological category and its associated morphological rule, therefore, are also described as more productive than a small category and its rule. The importance of productivity for studies of the lexicon and lexical processing is witnessed by the fact that no finite lexicon will suffice to process unseen text: Computational tools cannot do without taking productive word formation into account (see Article 31).

A first key question in productivity research is what conditions need to be met for a rule to be productive in these ways. A second key question is whether a rule is ever totally unproductive, i.e., whether productivity is in essence a graded phenomenon. A related issue is how the degree of productivity of a morphological category might be measured. A third set of questions addresses how productivity changes through time, and how affixes are used across registers by different social groups. A final issue is the relation between productivity and processing constraints in the mental lexicon.

How these questions are answered, and the relevance of corpora for guiding research, however, depends crucially on what is viewed as the goal of morphological theory.

2. Theoretical frameworks

The goal of morphological theory as defined in early generative morphology is to account for what complex words are possible words. In this approach, productive and unproductive rules have very different properties. First, productive rules are seen as the true rules, they are dynamic, and crucial for producing and understanding the associated complex words. Unproductive rules, by contrast, are viewed as only describing the structure of complex words that have been committed to memory. Unproductive rules have been described as redundancy rules (Jackendoff 1975), redundant in the sense that they describe (and account for) existing structure but do not play an active rule in production or comprehension. Second, regularity is taken to be a necessary condition for productivity, but not a sufficient condition - there are unproductive morphological categories with fully regular (redundancy) rules (Dressler 2003). Third, morphological rules are viewed as part of an internally consistent module of the grammar that characterizes the knowledge of an ideal speaker in a homogeneous speech community. This knowledge of the ideal speaker is assumed to be an adequate characterization of the knowledge of actual speakers. By implication, what real speakers actually say must provide an imperfect window on their true morphological competence, a window that is distorted by intervening pragmatic, sociolinguistic and stylistic variables. Consequently, corpora are largely irrelevant. In the words of Dressler (2003, 54), " statistic approaches ... are of little relevance in itself, because they refer to language norm and to individual performances. In fact, all corpora data are performance data which

reflect the realisation of linguistic norms and thus only indirectly the realisation of the corpus producers' competence of the system of potentialities."

A series of findings challenges this classical view of morphology and productivity. The central role of gradedness in morphology that has emerged from a wide range of recent studies (see Hay/Baayen 2005, for a review) casts doubt on the usefulness of an absolute distinction between productive and unproductive rules. The productivity of probabilistic paradigmatic morphology invalidates the assumption that productivity crucially depends on regularity as described by straightforward syntagmatic rules, and reinforces the importance of schemas, constructions, and local generalizations (Albright/Hayes 2003; Baayen 2003; Bybee 2001; Dabrowska 2004). In this emerging new theory, morphological productivity can be understood as resulting from a great many factors such as the individual language user's experience with the words of her language, her phenomenal memory capacities, her conversational skills, her command of the stylistic registers available in her language community, her knowledge of other languages, her communicative needs, her personal language habits and those of the people with which she interacts.

There are many ways in which corpora contribute to making progress towards the highly ambitious goal of understanding morphological productivity in its full complexity. Corpora allow researchers to explore how productivity varies across registers, written versus spoken language, social and geographical space,

and even time. Corpus-derived measures play an increasingly important role in research on lexical processing in the mental lexicon, and have proved essential for developing rigorous and falsifiable models for processing constraints on productivity. A first step in this direction was provided by the development of statistical measures of productivity.

3. Measuring productivity

Several corpus-based measures are now available for gauging different aspects of the productivity (Baayen 1992,1993; Baayen/Renouf 1996).

3.1. Mathematical formalizations of productivity

A first measure of productivity focuses on the size of the morphological category. A category with many members is more productive in the sense that it has produced many complex words that are useful to the language community. A rule that is highly productive in this sense is like a successful company selling a product that has a large share of the market. Such a rule has a high REALIZED PRODUCTIVITY. Realized productivity is similar to profitability in the sense of Corbin (1987), see also Bauer (2001: 49), but restricted to 'past achievement'. In Baayen (1993), it is refered to as extent of use. The realized productivity of a morphological category C is estimated by the type count V(C,N) of its members in a corpus with N tokens.

A second measure of productivity assesses the rate at which a morphological category is expanding and attracting new members. A category that is expanding at a higher rate is more productive than a category that is expanding at a lower rate, or that is not expanding at all. A rule that is highly productive in this sense is like a company that is expanding on the market (independently of whether that company has or does not have a large share of the market). Such a rule has a high EXPANDING PRODUCTIVITY. Expanding productivity is similar to Corbin's profitability, but oriented to what is expected for the near future. This aspect of productivity is estimated by means of the number of words V(1,C,N) in morphological category C that occur only once in a corpus of N tokens, the hapax legomena. Let V(1,N) denote the total number of hapax legomena in the corpus. The ratio P* = V(1,C,N)/V(1,C) is an estimate of the contribution of morphological category C to the growth rate of the total vocabulary. This measure is refered to as the hapax-conditioned degree of productivity (Baayen 1993).

A company may have a large share of the market, but if there are hardly any prospective buyers left because the market is saturated, it is nevertheless in danger of going out of business. A third measure of productivity gauges the extent to which the market for a category is saturated. A rule with a low risk of saturation has greater potential for expansion, and hence a greater POTENTIAL PRODUCTIVITY. The potential productivity of a rule is estimated by its

hapax legomena in the corpus divided by the total number of its tokens N(C) in the corpus: P = V(1,C,N)/N(C). This ratio, known as the category-conditioned degree of productivity (Baayen 1993), estimates the growth rate of the vocabulary of the morphological category itself.

PLACE FIGURE 1 APPROXIMATELY HERE

Figure 1. The dynamics of vocabulary growth. The horizontal axis displays corpus size in tokens (N), the vertical axis displays the number of types observed as the corpus size is increased. Solid lines represent two growth curves. The extent of use is the highest point of a curve. Potential productivity is defined as the slope of the tangent to the curve at its endpoint (dashed line). The dotted line illustrates that potential productivity depends on N, and decreases with increasing N.

All three measures are defined with respect to the statistical properties of word frequency distributions (Baayen 2001). A corpus providing a synchronic sample of the language can be viewed as a text that is to be processed from beginning to end. For each successive word token read, we note the total number of different word types observed thus far. When the number of types is plotted against the number of tokens, curves such as shown in Figure 1 are obtained. The measure for potential productivity (P) represents the rate at which the vocabulary is increasing at the end of the curve. It is mathematically equivalent

to the slope of the tangent to the curve at its endpoint.

A corpus can also be viewed as a (simplified) model of diachrony, as through life, new samples of text are continuously added to one's cumulative experience. In this case, the statistical theory of vocabulary growth curves presents the simplest possible model of how past experience combines with expectations for the near future. However, since most corpora present synchronic slices of adult language use, they are not well suited for studying diachronic change, not for the individual speaker, nor for communities of speakers. In psycholinguistics, the absence of diachronic corpora representing language input from birth to old age is acutely felt, and has led to experimental measures gauging age of acquisition, beginning with the study by Carroll/White (1973). Estimates of age of acquisition often have superior predictivity for lexical processing compared to synchronic frequency counts, and suggest that corpora sampling speech from different stages will be important resources for (corpus) linguistics as well.

The growth rate as measured with the P statistic is based on probability theory. An alternative computational method for estimating the growth rate is based on the deleted estimation method of Jelinek/Mercer (1985). Nishimoto (2003) shows that similar productivity rankings are obtained when potential productivity is estimated with this technique.

All these productivity measures are based on a partition of the vocabulary into morphological categories. As a consequence, only those words in which a given rule was the last to apply are taken into account. For instance, HELPFULNESS is counted only once, namely, as a member of the morphological category of words in NESS. It is not counted as a member of the category of words in FUL. If one were to assign HELPFULNESS to both morphological categories, the observations in the two categories would no longer be independent, and the statistical tests for comparing degrees of productivity would no longer be valid. On the other hand, it certainly does make sense to consider HELPFULNESS as instantiating a particular use of a member of the morphological category of FUL. When lack of statistical independence is not a problem, for instance, when the focus of interest is on ranking affixes by their degree of productivity, words with a given affix that fall outside of the corresponding morphological category proper can be taken into account as well. Gaeta/Ricca (2005) have shown that similar rankings are obtained for counts excluding and counts including words that have undergone further word formation. This suggests that measures that are based strictly on the morphological category itself provide a good approximation that has, on the one hand, the advantage of ease of extraction from (unanalyzed) corpora, and on the other hand, the advantage of allowing further statistical testing.

3.2. Interpretation and validation

These three productivity measures are statistical formalizations of the intuitive notion of productivity. To what extent are these measures linguistically interpretable, and to what extent have they been validated?

3.2.1. Realized productivity

First consider the simple type count that provides a corpus-based estimate of realized productivity. According to Bybee (2001), the productivity of a word formation schema is largely determined by its type frequency. For instance, the English past tense in ED is realized on thousands of verbs, whereas irregular schemas such as that exemplified by KEEP/KEPT and SLEEP/SLEPT, pertain to only small numbers of verbs. The regular past tense schema has a much greater realized productivity than the irregular past tense schemata. The importance of type frequency emerges even more clearly when it is pitted against token frequency. Productive categories are characterized by the presence of large numbers of low-frequency forms, whereas unproductive categories tend to contain many high-frequency forms, unsurprisingly, as a high token frequency protects irregular forms against regularization and helps explain their continued existence. Assessing the productivity of a schema in terms of token frequency would therefore be counterproductive. For instance, Baayen/Moscoso del Prado Martin (2005) studied 1600 monomorphemic English verbs, of which 146 were irregular and 1454 regular. The summed frequency of all irregular verbs, 1793949, exceeds the summed frequency of the much larger set of regular verbs (732552) by a factor 2.5 (counts based on Baayen/Piepenbrock/Gulikers 2005, which is based on a corpus of 18.5 million tokens). An assessment of the productivity of the past tense in English using token frequency would suggest that vocalic alternation would be more

productive than suffixation with ED, contrary to fact.

Type frequency, however, provides only a first approximation of the productivity of a schema. First of all, it does not take into account the degrees of similarity between the words that are governed by that schema. Much improved estimates of the productivity of a particular schema are obtained with analogical models (e.g., Skousen 1989; Albright/Hayes 2003). Second, type-based counts do not do justice to the lower weight of low-frequency words that speakers may not know well or not know at all. In this respect, analogical models also offer further precision (see Skousen 1989, for token-weighted analogy). Third, type-based counts work reasonably well in a fixed domain like past tense inflection in English, where morphological alternatives carrying the same function are compared. Type-based counts fare less well, however, when very different morphological categories are compared. For instance, Dutch has several suffixes for creating nouns denoting female agents. The most productive of these suffixes is STER, as in VERPLEEG-STER, 'female nurse' (compare VERPLEEG-ER, 'male nurse'). Dutch also has a verb-forming prefix VER (as in VER-PLEEG-EN, 'to nurse'), that is described as less productive. Even though STER is judged intuitively to be more productive than VER, the type frequency of VER (985) is much higher than the type frequency of STER (370) in counts based on a corpus of 42 million words. This shows that there are aspects of productivity that are not well-represented by a category's realized productivity. A morphological category may have a high realized productivity, but a

high realized productivity does not imply that its expanding productivity or its potential productivity will be high as well.

3.2.2. Expanding productivity

The relative rate at which a category is expanding provides a first complement to the type-based estimate of productivity. This measure generates productivity rankings that provide reasonable reflections of linguists' overall intuitions about degrees of productivity (e.g., Baayen 1993). Recall that the degree to which a category is expanding can be estimated by considering that category's contribution to the growth rate of the vocabulary in a corpus. In practise, differences in expanding productivity can be gauged simply by comparing counts of hapax legomena.

Gaeta/Ricca (2005) propose an alternative measure that is also based on the count of hapax legomena. They argue that counts of hapax legomena should be compared when equal numbers of tokens have been sampled for each of the morphological categories involved. With reference to Figure 1, they propose to measure the growth rate for the endpoint of the short curve and for the same number of tokens for the upper curve. This amounts to comparing the slopes of the lower dashed line and the upper dotted line. It turns out that their 'variable corpus' productivity measure is mathematically and empirically closely related to the present measure for expanding productivity. Both lead to plausi-

ble productivity rankings of derivational categories in Italian.

Hapax legomena should not be confused with neologisms. An ideal lexicographic measure of expanding productivity would specify the rate at which neologisms with a given morphological structure are added to the vocabulary of the language community, calibrated with respect to the rate at which the community's vocabulary as a whole is expanding. For the problems involved with approaches based on neologisms, see section 4.1.2. Corpus-based counts of hapax legomena provide an indirect way of estimating the rate at which a morphological category enriches the vocabulary. For small corpora, the number of neologisms among the hapax legomena will be small. As the corpus size increases, the number of neologisms in the corpus increases as well. Crucially, these neologisms are found primarily among the hapax legomena, and to a lesser extent among the words occurring twice, three times, etc. (Baayen/Renouf 1996; Plag 2003). Nevertheless, even in a large corpus, there may be words among the hapax legomena that have been well-established in the language for centuries. This is not a problem as long as it is kept in mind that the hapax legomena are not a goal in itself, they only function as a tool for a statistical estimation method aimed at gauging the rate of expansion of morphological categories.

The degree to which a morphological category is expanding captures an important aspect of what it is to be productive. But there is a further aspect of

productivity that is not properly captured by the count of types nor by the count of hapaxes. Recall that the Dutch suffix STER is in some sense more productive than the Dutch prefix VER. A type-based comparison failed to bring this difference to light, and a count of hapax legomena (274 for VER, 161 for STER) fails to do so as well. The problem with these two counts is that they do not do justice to the intuition that it is easy to think of new well-formed words in STER, but very hard to think of well-formed neologisms in VER.

3.2.3. Potential productivity

The ratio of hapax legomena in a given morphological category to the total number of tokens in that category, the category's potential productivity, assigns VER a productivity index of 0.001 and STER a much higher index of 0.031. This measure for a category's potential productivity indicates correctly that it is easier to think of a neologism in STER than of a neologism in VER.

The potential productivity measure is highly sensitive to markedness relations. The unmarked suffix for creating agent nouns in Dutch is ER (GEEF-ER, 'giver'), its marked counterpart is STER (GEEF-STER, 'female giver'). Unmarked ER has the greater realized productivity as well as the greater expanding productivity, but marked STER has the greater potential productivity.

What the potential productivity measure highlights is that productivity can be

a self-defeating process, in the sense that once an affix has saturated the onomasiological market, it has no potential for further expansion. Unmarked ER has saturated its market to a much greater extent than has STER. As a consequence, STER can freely attach to a great many verbs where it has not been used before, due to the reluctance of Dutch speakers to explicitly mark the gender of agents.

An experimental study validating the potential productivity measure is Baayen (1994b). Following Anshen/Aronoff (1988), subjects were asked to generate within 5 minutes as many words with a specified affix as they could think of. Exactly as predicted, subjects produced many more neologisms in STER than in ER or VER. Further validation of this measure has been provided by Hay (2003), who showed that it is correlated with measures for the parseability of the complex words in the morphological category (see section 4.2.2.). Furthermore, Wurm/Aycock/Baayen (2006) observed that the potential productivity measure is predictive for visual lexical decision latencies.

The potential productivity measure is also sensitive to the compositionality of the words in the morphological category, albeit indirectly. Words with less compositional meanings typically tend to be high-frequency words. Since the token frequencies of the words in the morphological category contribute to the denominator of the potential productivity measure, the presence of opaque words will tend to lead to lower estimates of potential productivity.

A closely related measure for potential productivity is the ratio I of the estimated size of the category S in an infinitely large corpus and the observed number of types in a corpus of size N: I = S/V(N). This ratio quantifies the extent to which the attested types exhaust the possible types. Rough estimates of S are provided by statistical models for word frequency distributions, such as the finite Zipf-Mandelbrot model developed in Evert (2004). Affixes with a high potential productivity also tend to have a high I (see, e.g., Baayen, 1994b), indicating that many more types could be formed than are actually attested in the corpus.

The validity of these productivity measures hinges on the availability of correct input data. Generally, string-based searches of affixes in corpora produce highly polluted word frequency distributions that may seriously distort the true pattern in the data. Manual inspection and correction, although time-consuming, is as crucial for productivity research (Evert/Lüdeling 2001) as it is for research on syntax (see Article 45).

4. Forces shaping productivity

Traditional approaches to morphological productivity have invested in finding structural explanations for degrees of productivity. One influential view originating from early structuralism (see Schultink 1961) is that the degree of productivity is inversely proportional to the number of grammatical restrictions on that rule. However, it is difficult to see how the quantitative effects of

the very different kinds of structural constraints would have to be weighted. As pointed out by Bauer (2001: 143), "words are only formed as and when there is a need for them, and such a need cannot be reduced to formal terms".

If structural constraints as such do not directly drive morphological productivity (see section 4.2.3. for indirect effects, however), research on morphological productivity should be directed towards other factors. There are two clusters of such factors that play a demonstrable role, one cluster pertaining to societal factors, the other to the role of processing constraints in the mental lexicons of individual speakers.

4.1. Productivity in the speech community

It is well-known that there is consistent variation in how speakers with different backgrounds and varying communicative goals make use of the morphological and grammatical constructions offered by their language. Biber (1988), for instance, provided detailed evidence that the linguistic resources employed in speech differ from those in writing, and that within each of these communicative modalities, further systematic differences differentiate the styles of more specific text types. Contemporary work in stylometry (e.g., Burrows 1992) showed, furthermore, that individual writers develop their own characteristic speech habits, not only in the selection of their topics, but, crucially, in which grammatical resources offered by the language authors typically tend to

use (see also Articles 40 and 52).

This work in corpus linguistics has had little impact on productivity research in theoretical morphology. Bauer's monograph (2001) reveals no awareness of the possibility that morphological categories might be more productive in some registers than in others, and the potential consequences of such stylistic forces for the weight of structural constraints in explanations of productivity. However, the little work that has been done in this area shows unambiguously that, unsurprisingly, different genres recruit different morphological categories to very different degrees.

A further complication in productivity research is that the needs of speech communities and groups of specialists within these speech communities (Clark, 1998) change over time. In modern technological societies, the ever increasing rate of scientific and technological progress leads to a proliferation of new techniques, concepts and products that require names. How productive affixes are used, and the rate at which new words appear through the years will depend on whether discourse is studied from a domain with rapid innovation or with slow or little innovation. In what follows, we first consider productivity in relation to register variation. Next, we consider productivity from the perspective of the society and its changing needs.

4.1.1. Register and productivity

Plag/Dalton-Puffer/Baayen (1998) showed, using the British National Corpus, that the degree of productivity of a suffix may differ depending on whether it is used in written language, formal spoken language, or informal spoken language. Most derivational suffixes were observed to be more productive in written than in spoken language, with the exception of WISE. Furthermore, the productivity rankings of affixes changed from one register to the other. For instance, NESS emerged as more productive than ABLE in written language, but in spontaneous conversations, ABLE was slightly more productive than NESS. The different degrees of productivity observed for spoken and written language are expected given the very different conditions under which oral and written language are produced. Oral language tends to be produced on the fly, written language tends to go through several rounds of revision before it appears in print. Furthermore, oral language is anchored in the physical context, where prosody, gesture, gaze, and common ground provide very different constraints on communication compared to written language, where sentence and discourse structure have to bear the full burden of communication. Thus, the greater productivity of NESS in written discourse observed by Plag/Dalton-Puffer/Baayen (1998) may be due to the possibility to use this suffix to refer to states of affairs previously introduced into the discourse (Kastovsky 1986; Baayen/Neijt 1997).

PLACE FIGURE 2 APPROXIMATELY HERE

Figure 2. Selected affixes in the space spanned by the first two dimensions resulting from a principal components analysis of the correlational structure of their potential productivity in four different kinds of texts (stories for children, officialese, literary texts, and religious texts). Individual texts are shown in grey (after Baayen 1994a).

Different registers tend to be used for communication about very different kinds of topics. The suffix ITY, for instance, is more productive in scientific and technical discourse, and the suffix ITIS appears predominantly in medical discourse and occasionaly in non-medical texts in certain journalistic registers (Lüdeling/Evert 2005, see also Clark 1998). A study addressing differences in potential productivity accross various registers of written English using methods from stylometry is Baayen (1994a). Figure 2 visualizes the correlational structure for selected affixes and texts using principal components analysis. Germanic affixes are found predominantly in the right half of the plot, Latinate affixes occur more to the left. The texts that have a preference for the Germanic affixes are, for instance, the stories for children by Lewis Carroll and Frank Baum. The Latinate affixes, by contrast, are most productive in officialese, Startrek novels, and the scientific proze of William James. Register variation challenges theoretical approaches that seek to ground the productivity of morphological categories purely in structural constraints, since these constraints do not vary with register. Register variation was also observed by Baayen/Neijt

(1997) for the Dutch suffix HEID, the translation equivalent of English NESS. They studied a newspaper corpus, and found that HEID was not productive at all the articles on economics, and most productive in the sections on literature and art.

4.1.2. Productivity through time

Languages change as time passes by. New morphological categories may come into existence and established categories may fade away, while other categories always remain peripheral, drifting along the tides of fashion (e.g., non-medical ITIS, Lüdeling/Evert 2005, or COD and FAUX in British English, Renouf/Baayen 1997).

Traditionally, the diachronic aspects of productivity have been studied by means of dictionaries. Especially the Oxford English Dictionary has proved to be a useful source of information, as it provides dates for first and last mentions (Neuhaus, 1973, Anshen/Aronoff 1999, Bolozky 1999). However, the use of dictionaries brings along several methodological problems. Especially for the older stages of the language, the sampling is - unavoidably - sparse. As a consequence, a word may have been in use long before it is first observed in the historical record. Conversely, for recent developments, the sheer volume of text available both in print and on the internet prohibits exhaustive description. Furthermore, dictionaries provide little control over variation in produc-

tivity due to register. Finally, words with a new onomasiological function are of lexicographical interest and are relatively easy to detect, while words with referential functions (in the sense of Kastovsky 1986) tend to go unnoticed.

Diachronic studies of language change, however, need not be based on dictionaries (see chapter 6). Ellegard (1953), for instance, is a classic example of a corpus-based study long before electronic corpora were available. The diachronic study of German nominal ER by Meibauer/Guttropf/Scherer (2004) is similar in spirit. They carefully extracted samples from four centuries of German newspapers, which were digitized and analyzed semi-automatically. This procedure is exemplary, in that it controls to a large extent for register variation.

In general, the use of corpora sampling texts at different points in time is not without its own share of methodological pitfalls, unfortunately. Data sparseness remains a problem, especially for the older stages of the language. Furthermore, any imbalance in the materials included invariably leads to diachronic artefacts, such as the sudden increase in the use of medical ITIS in the first decade of the 20th century observed by Lüdeling/Evert (2005), which they were able to trace to the inclusion of an encyclopedia published in 1906.

What is clear from diachronic studies of productivity is that probabilistic models that assume a fixed population of possible words from which successively more words are sampled as time proceeds are fundamentally flawed. Studies

based on the OED show bursts of productivity around 1600 and around 1850 (Neuhaus 1973; Anshen/Aronoff 1999), but changes may even be going on at much shorter time scales of just a few years (Baayen/Renouf 1996; see also chapter 64 for corpora in the study of recent language change).

The study of Meibauer/Guttropf/Scherer (2004) offers an interesting perspective on the increasing realized and potential productivity of nouns in ER in German newspapers. Through time, ER is attached more often to complex base words, it is increasingly more productive as a deverbal suffix, and it primary function has become to denote persons. Meibauer and colleagues point out that a similar development characterizes the micro-level of children acquiring German, and they offer several explanations of why this similarity might arise. Possibly, the linguistic development reflects, at least in part, the expanding cognitive, social and intellectual skils of both the child and of the increasingly complex society of speakers of German as it developed over 400 years.

The historical record also reveals that morphological categories may cease to be productive. Anshen/Aronoff (1997, 1999) discuss English OF and AT (which dropped out of use within a short period of time) and MENT (which showed a much more gradual decline). Keune/Ernestus/Van Hout/Baayen (2005) provide further detail on the decline of productivity using a corpus of spoken Dutch. They show how loss of productivity is reflected in the reduction of the acoustic realizations of the highest-frequency members of the category. Words

such as NATUUR-LIJK (literally 'nature-like', but with the opaque meaning 'of course') can be reduced to TUUK in spontaneous conversations, which shows these words are in the process of becoming monomorphemic.

4.2. Productivity and processing constraints in the mental lexicon

Productivity is subject not only to societal forces, but also to cognitive constraints governing lexical processing in the mental lexicon of the individual. Corpus-based surveys of actual use and corpus-based estimates of a wide range of lexical, sublexical, and supralexical probabilities play a crucial role in psycholinguistic research on productivity.

4.2.1. Productive and unproductive: an absolute distinction?

Many researchers view productivity as a diagnostic that can be used "to determine which patterns are fossilized, and which represent viable schemas accessible to speakers" (Bybee 2001:13). The implication is that productive morphology would be in some sense cognitively more real than unproductive morphology, and that it would make sense to make a principled distinction between being productive (to a greater or lesser degree) and being totally unproductive. Recent results in mental lexicon research argue against such an absolute split between live rules and fossilized residues.

First, it has become clear that complex words leave traces in lexical memory, irrespective of whether they are regular or irregular (see Hay/Baayen 2005). The frequency with which a complex word is used may even co-determine the fine acoustic detail of its constituents (Pluymaekers/Ernestus/Baayen 2005). Consequently, a distinction between totally unproductive rules or schemas on the one hand, and productive rules or schemas on the other hand, would imply that the stored exemplars of unproductive schemas would not be available for generalization, while the stored exemplars of productive rules would allow generalization. This is not only implausible, but also contradicts the well-documented finding that the schemas of irregular verb classes can serve as attractors in both synchrony and diachrony (Bybee/Slobin 1982).

Second, the same kind of gang effects that characterize the unproductive schemas for the irregular past tense in English have been shown to be active for the regular past tense in ED as well (Albright/Hayes 2003). The strength of such gang effects, and not regularity or default status as such, has also been shown to predict the productivity of case inflection in Polish (Dabrowska, 2004). Given that exactly the same analogical mechanisms underly both the irregular and the regular past tense, the difference between the unproductive irregular and the productive regular forms is a matter of degree and not a matter of fundamentally different cognitive principles (see also Pothos 1985, but Pinker 1997; Anshen/Aronoff 1999; Dressler 2003, for the opposite position).

It might be argued that the 'semi-productivity' of certain irregular inflections contrasts with the complete lack of productivity for a derivational suffix like TH in English WARMTH and STRENGTH, which is the textbook example of an unproductive suffix (e.g., Bauer 2001:206). Nevertheless, new words in TH are occasionaly used, as illustrated by the following text: "The combination of high-altitude and low-latitude gives Harare high diurnal temperature swings (hot days and cool nights). The team developed a strategy to capture night-time coolth and store it for release the following day. This is achieved by blowing night air over thermal mass stored below the verandah's ..." (http://www.arup.comp/insite/features/printpages/harare.htm, observed in 2001). When considered out of context, COOLTH seems odd, jocular, perhaps literary or even pretentious. Yet inspection of how it is actually used in context reveals that COOLTH fills an onomasiological niche by supplying a word denoting a quantity of the physical property referenced by the adjective COOL that fits with the existing words in TH that express similar quantities (WARMTH, STRENGTH, LENGTH, and WIDTH). COOLTH is a possible word of English, but it is at the same time a very low-probability word of English, not because speakers of English are unaware of the structural similarity of WARMTH, STRENGTH, LENGTH, and WIDTH, but because the probability that TH can be used to satisfy a sensible onomasiological need is extremely low. For a similar example from Dutch, see Keune/Ernestus/Van Hout/Baayen (2005).

4.2.2. Processing constraints

We have seen that productivity is co-determined by register as well as by the onomasiological needs in a language community, and that the same cognitive principles underlie both unproductive and productive rules: exemplar-driven analogical generalization. Productivity is further restricted by processing constraints.

Recall the prominence of the count of hapax legomena in the measures for expanding and potential productivity. From a processing perspective, the hapax legomena represent the formations with the weakest traces in lexical memory. Consequently, it is for these words that comprehension and production is most likely to benefit from rule-driven processes. Conversely, when a morphological category comprises predominantly high-frequency words, strong memory traces for these words exist that decrease the functional load for production and comprehension through rule-driven processes. Hence, the importance of rules for the lexical processing of complex words will be larger for morphological categories with many low-frequency words.

There are two ways in which the consequences of lexical processing for productivity can be made more precise, as shown by Hay (2003).

First, the frequency of the base relative to that of the derived word should be

taken into account. Many complex words have a lower frequency than their base word (e.g., ILLIBERAL). But there are also complex words that are more frequent than their bases (e.g., ILLEGIBLE). The greater the frequency of the complex word compared to that of its base, the greater the likelihood will be that its own memory trace will play a role during lexical access. Conversely, rule-driven processing will be more important for formations with memory traces that are much weaker than those of their constituents.

Effects of relative frequency (defined here as the frequency of the base divided by the frequency of the derivative) have been observed both in production and in comprehension. In English, t-deletion is more likely for words with a low relative frequency such as SWIFTLY (221/268) than for words with a high relative frequency such as SOFTLY (1464/440) (Hay 2001). SWIF(T)LY, in other words, is in the process of becoming independent of its base word, its simplified phonotactics indicate it is becoming more like a monomorphemic word in speech production.

In comprehension, relative frequency is an indicator of the relative importance of decompositional processing. As shown in Hay/Baayen (2002), morphological processing in reading can be understood in terms of lexical competition between the base and the derivative, such that the derivative (the whole) has a small headstart over the base (one of its parts). In their model, words with a high relative frequency are accessed primarily through their base, while words

with a low relative frequency are accessed primarily on the basis of their own memory traces.

PLACE FIGURE 3 APPROXIMATELY HERE

Figure 3. Log potential productivity as a function of the proportion of types that are parsed in the model of Hay/Baayen (2002).

It turns out that relative frequency predicts potential productivity. Hay (2003) showed that the proportion of types in a morphological category for which the base frequency exceeds the frequency of the derivative is positively correlated with the logarithmic transform of its potential productivity (P). Hay/Baayen (2002) showed that estimates of the proportion of types that are parsed (according to a computational model of morphological processing in reading) likewise predicts potential productivity (see Figure 3). The advantage of this modeling approach is that the perceptual advantage of the whole over its parts (see also Hay/Baayen 2005) as well as the frequency and length of the affix are taken into account. It is remarkable that 41% of the variance in the (logged) P measure of 80 English morphological categories can be accounted for by relative frequency alone. One way of interpreting relative frequency as a predictor of productivity is that it gauges the experience with the rule in production and comprehension. A greater relative frequency implies stronger memory traces for the rule itself, and hence an intrinsically increased potential for producing

and understanding new words.

A second important processing constraint concerns the derivative's junctural phonotactics, the probability of the sequence of sounds spanning the juncture between its parts. Low-probability sequences such as NH in INHUMANE create boundaries that are highly unlikely to occur within morphemes, and hence provide probabilistic information about morphological complexity. For speech production, low-probability sequences require more articulatory planning whereas high-probability sequences may benefit from automatized gestural scores. Hence, an affix is more likely to be an independent unit in speech production if it tends to create low-probability junctural phonotactics. In comprehension, the constituents are easier to parse out for words with low-probability junctures; for experimental evidence see Seidenberg (1987), Hay (2003) and Bertram (2004). Given that complex words with low-probability junctures are easier to parse, affixes that create words with low-probability junctures should be more productive. Hay (2003) and Hay/Baayen (2003) observed just this: Several measures of junctural phonotactics (derived from the token frequencies of 11383 English monomorphemic words in a corpus of 18 million words) correlated signifantly with all three abovementioned measures of productivity. For instance, the junctural probability averaged over all words in the morphological category explained some 14% of the variance in (log) potential productivity.

4.2.3. Conspiracies

Relative frequency and junctural phonotactics are involved in two correlational conspiracies.

The first conspiracy concerns the strong intercorrelations of all measures of productivity, measures for junctural phonotactics, measures for relative frequency and parsing, and lexical statistical measures such as Shannon's Entropy. Hay/Baayen (2003) observed, using principal components analysis, that this correlational structure has two orthogonal dimensions of variation. The first dimension represents the tight intercorrelations between measures that guage how affixes are used against the backdrop of the corpus as anchor point for normalisation. Realized productivity, expanding productivity, entropy (information load), the count of formations with low-probability junctural phonotactics, and the estimates of the number of types parsed all enter into strong positive correlations. These measures quantify aspects of the past and present usefulness of an affix. This dimension of variation is probably most closely linked to its onomasiological usefulness in society, its referential functionality (Kastovsky 1986; Baayen/Neijt 1997), and register variation.

The second dimension unifies measures that are normalized with respect to the individual morphological categories. Potential productivity, the estimated proportion of types in the category that are parsed, and the frequency of the base

(averaged over the types in the category) enter into strong positive correlations, and reveal strong negative correlations with the frequency of the derivative (averaged over the types in the category) and with the probability of the juncture (similarly averaged). This dimension guages the strength of the rule in terms of the proportion of words in the corresponding morphological category that are accessed in comprehension and production primarily through that rule rather than through the memory traces of the derivatives themselves.

PLACE FIGURE 4 APPROXIMATELY HERE

Figure 4. Log potential productivity as predictor of complexity-based rank. The two words marked with an asterisk exchange rank in structure-based ordering. Affixes in grey were identified as outliers and were excluded when calculating the regression line (estimated slope: 1.95, R-squared: 0.47).

The second conspiracy involves processing constraints, grammatical constraints, and memory constraints.

Hay/Plag (2004) showed that English suffixes can be arranged in a hierarchy such that their rank in the hierarchy is predictive for the order in which these suffixes can occur in complex words. Given that a suffix has rank i, suffixes with rank greater than i may follow that suffix in a word, while suffixes with rank lower than i will never follow it. The position of a suffix in this hierarchy is predictable from measures gauging the strength of the suffix such as potential productivity and the proportion of types in the morphological category that are parsed (the second dimension identified above for the first conspiracy). This is illustrated in Figure 4 (based on Table III of Hay/Play, 2004) for potential productivity. As the log-transformed potential productivity increases, the word's rank increases as well. Hay (2003) argues that this hierarchy is driven by processing complexity: An affix that can be easily parsed out should not precede an affix that is more difficult to process. Hay/Plag (2004) refer to this

as the hypothesis of complexity-based ordering. In other words, suffixes that are more productive, and that function more as processing units in speech production and comprehension, must follow less productive suffixes.

The hypothesis of complexity-based ordering raises two questions. The first question concerns the goal of word-formation: the creation of new words for communication. Is this goal completely subordinated to low-level processing constraints? This unlikely possibility has been ruled out by Hay/Plag (2004), who showed that grammatical restrictions (such as the required syntactic, semantic and phonological properties of the base) lead to a nearly identical hierarchy as that established on the basis of attested suffix combinations. This hierarchy can also be predicted from potential productivity and related processing measures. What seems to be at stake, then, is a conspiracy of on the one hand potential productivity and its correlated processing constraints, and on the other hand grammatical constraints, all of which work in tandem to optimize the structure of multiply suffixed words for communication.

The second question prompted by the hypothesis of complexity-based ordering is why comprehension would benefit from less productive affixes preceding productive affixes. After all, there is no evidence that parsability constrains possible sequences of words in sentences in a similar way.

As a first step towards an answer, we note that this conspiracy of processing

and grammatical factors is reminiscent of another probabilistic ordering hierarchy first discussed by Bybee (1985) for inflection. Bybee showed that more inherent inflection (tense and aspect marking, for instance) tends to be realized closer to the stem or root than more contextual inflection (person and number marking, for instance). In this inflectional hierarchy, therefore, the more (formally and semantically) predictable formatives are found to be peripheral to the less predictable and more fusional exponents, just as in derivation the less productive and often semantically less predictable suffixes are closer to the stem. What this suggests is that semantic transparency is also at issue. Especially the syntactic and semantic grammatical constraints studied by Hay/Plag (2004), which guarantee a minimal level of transparency, point in this direction.

A further step towards an answer can be made by a more careful consideration of the role of frequency in morphological processing. Hay/Baayen (2002), following Hay (2003), assume that relative frequency primarily affects low-level processes at the level of form. However, word frequency is more strongly correlated with measures of a word's meaning than with measures of a word's form (Baayen/Feldman/Schreuder 2006), and it seems likely that a measure such as relative frequency (and derived measures such as parsing ratios) also reflects the relative complexities of compositional processes at the level of semantics. This may help explain why a conspiracy of processing constraints and grammatical constraints can exist.

Finally, it has been observed that productivity is inversely proportional to the likelihood of serving as input for further word formation. More specifically, Krott/Schreuder/Baayen (1999) reported that a greater potential productivity enters into a negative correlation with the proportion of the derivatives in the morphological category that serve as input to further word formation. Frequent, short words with less productive affixes are more likely to produce morphological offspring than more ephemeral complex words. This is exactly as expected from an onomasiological perspective, as well as from a processing perspective: More frequent words are more readily available in lexical memory as input for not only syntactic but also for morphological processing. This is, of course, the other side of the coin of complexity based ordering, but it extends beyond affix ordering to derivatives in compounds.

To conclude, potential productivity is part of a correlational conspiracy of different factors: low-level perceptual factors (as evidenced by junctural phonotactics and relative frequency), factors pertaining to morphological processing at the levels of form and meaning (as evidenced by relative frequency and selectional restrictions), and factors arising at the interface of memory and onomasiological needs.

5. Concluding remarks

What, then, is morphological productivity? Many theoretical morphologists

have attempted to define productivity as a property of the language system (e.g., Schultink 1961; Bauer 2001; Dressler 2003). Unfortunately, these definitions and the underlying theories have not led to models with predictive power for degrees of productivity. At the same time, traditional research has been dismissive of the potential relevance of system-external factors. Bauer's definition of productivity (2001:211) is telling, in that it states that the extent to which a morphological category is actually used "may be subject unpredictably to extra-systemic factors". Contrary to what Bauer suggests, recent research has shown not only that the effects of 'extra-systemic' factors are truly predictive for productivity, but also that the 'intra-systemic' factors are part of a much larger system of interacting factors.

Exciting new insights have been obtained precisely by combining historical, stylistic, onomasiological, and cognitive factors in a quantitative and hence falsifiable empirical research paradigm. Without corpora, these insights would never have been obtained, prediction — the goal of scientific inquiry — would have remained out of reach, and productivity research would never have emerged from the quagmire of studies providing overviews and syntheses of previous studies based on idiosyncratic, small, and non-representative data. However, much still remains to be done, even new fields await exploration, such as the role of sociolinguistic variables or the role of word formation in communal lexicons (Clark 1998). In short, in order to come to a full understanding of the challenging phenomenon of morphological productivity, a truly interdisci-

plinary data-driven research effort is required.

6. Literature (a selection)

Albright, Adam/Hayes, Bruce (2003), Rules vs. Analogy in English Past Tenses: A Computational/Experimental Study. In: "Cognition" 90, 119-161.

Anshen, Frank/Aronoff, Mark (1981), Morphological productivity and morphological transparency. In: "The Canadian Journal of Linguistics" 26, 63-72.

Anshen, Frank/Aronoff, Mark (1988), Producing morphologically complex words. In: "Linguistics" 26, 641-655.

Anshen, Frank/Aronoff, Mark (1997), Morphology in real time. In: Geert. E. Booij & Jaap van Marle (eds), "Yearbook of Morphology 1996". Dordrecht: Kluwer Academic Publishers, 9-12.

Anshen, Frank/Aronoff, Mark (1999), Using dictionaries to study the mental lexicon. In: "Brain and Language" 68, 16-26.

Baayen, R. Harald (1992), Quantitative aspects of morphological productivity. In: Geert. E. Booij & Jaap. van Marle (eds), "Yearbook of Morphology 1991", Dordrecht: Kluwer Academic Publishers, 109-149.

Baayen, R. Harald (1993), On frequency, transparency, and productivity. In: Geert E. Booij and Jaap van Marle (eds), "Yearbook of Morphology 1992". Dordrecht: Kluwer Academic Publishers, 181-208.

Baayen, R. Harald (1994a), Derivational productivity and text typology. In: "Journal of Quantitative Linguistics" 1, 16-34.

Baayen, R. Harald (1994b), Productivity in language production. In: "Language and Cognitive Processes" 9, 447-469.

Baayen, R. Harald/Neijt, Anneke (1997), Productivity in context: a case study of a Dutch suffix, "Linguistics" 35, 565-587.

Baayen, R. Harald/Renouf, Antoinette (1996), Chronicling The Times: Productive lexical innovations in an English newspaper. In: "Language" 72, 69-96.

Baayen, R. Harald (2001), "Word frequency distributions". Dordrecht: Kluwer Academic Publishers.

Baayen, R. Harald (2003), Probabilistic approaches to morphology. In: Bod, Rens, Hay, Jennifer B. & Jannedy, Stefanie, "Probability theory in linguistics", Cambridge: The MIT Press, 229–287.

Baayen, R. Harald/Feldman, Laura/Schreuder, Robert (2006), Morphological influences on the recognition of monosyllabic monomorphemic words. In: Journal of Memory and Language, in press.

Baayen, R. Harald/Piepenbrock, Richard/Gulikers, Leon (1995), The CELEX lexical database (CD-ROM), University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

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

Bertram, Raymond/Hyona, Jukka/Pollatsek, Alexander (2004), Morphological parsing and the use of segmentation cues in reading Finnish compounds. In: Journal of Memory and Language" 51, 325-345.

Biber, Douglas (1988), "Variation across speech and writing". Cambridge: Cambridge University Press.

Bolozky, Shmuel (1999), "Measuring productivity in word formation". Leiden: Brill.

Burrows, John F. (1992), Computers and the Study of Literature. In: C.S. Butler, "Computers and Written Texts". Oxford: Blackwell, 167-204.

Bybee, Joan L. and Slobin, Dan I. (1982), Rules and schemas in the development and use of the English past tense. In: "Language" 58, 265-289.

Bybee, Joan L. (2001), Phonology and language use, Cambridge: Cambridge University Press.

Carroll, John. B. and White, M. N. (1973), Age of Acquisition Norms for 220 Picturable Nouns. In: "Journal of Verbal Learning and Verbal Behavior" 12, 563-576.

Clark, Herbert H. (1998), Communal lexicons. In: K. Malmkjaer and J. Williams (eds), "Context in language learning and language understanding". Cambridge: Cambridge University Press, 63-87.

Corbin, Danielle (1987), "Morphologie derivationelle et structuration du lexique". Tübingen: Niemeyer.

Cowie, Claire (2003), Uncommon terminations: proscription and morphological productivity. In: "Italian Journal of Linguistics" 15, 99-130.

Dabrowska, Ewa (2004), Rules or schemas? Evidence from Polish. In: "Language and Cognitive Processes" 19, 225-271.

Dalton-Puffer, Cristiane/Cowie, Claire (2002), Diachronic word-formation and studying changes in productivity over time. Theoretical and methodolgical considerations. In: Javier E. Diaz Vera (ed) "A changing world of words. Studies in the English historical lexicography, lexicology and semantics (Consterus New Series 141)". Amsterdam, New York: Rodopi, 410-437.

Dressler, W. (2003), Degrees of grammatical productivity in inflectional morphology. In: "Italian Journal of Linguistics" 15, 31-62.

Ellegard, A. (1953), "The auxiliary do: The establishment and regulation of its use in English". Stockholm: Almquist & Wiksell.

Evert, Stefan/Ldeling, Anke (2001), Measuring morphological productivity: Is automatic preprocessing sufficient? In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie, & Shereen Khoja (eds.), Proceedings of the Corpus Linguistics 2001 conference, Lancaster, pp. 167–175.

Evert, Stefan (2004), A simple LNRE model for random character sequences. In: Purnelle, G. and Fairon, C. and Dister, A. (eds), "Le poids des mots. Proceedings of the 7th international conference on textual data statistical analysis". Louvain-la-Neuve: UCL, 411-422.

Gaeta, Davide/Ricca, Livio (2005), Productivity in Italian word formation: a variable corpus approach. In: "Linguistics" (to appear).

Hay, Jennifer B. (2001), Lexical frequency in morphology: Is everything relative?, In: "Linguistics" 39, 1041-1070.

Hay, Jennifer B. (2003), "Causes and Consequences of Word Structure". New York: Routledge.

Hay, Jennifer B./Baayen, R. Harald (2002), Parsing and productivity. In: Geert E. Booij & Jaap. van Marle (eds), "Yearbook of Morphology 2001", Dordrecht: Kluwer Academic Publishers, 203-235.

Hay, Jennifer B./Baayen, R. Harald (2004), Phonotactics, Parsing and Productivity. In: "Italian Journal of Linguistics", 15, 99-130.

Hay, Jennifer B./Baayen, R. Harald (2005), Shifting paradigms: gradient structure in morphology. In: "Trends in Cognitive Sciences", in press.

Hay, Jennifer B./Plag, I. (2004), What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. In: "Natural Language and Linguistic Theory", 22, 565-596.

Jackendoff, Ray S. (1975), Morphological and Semantic Regularities in the Lexicon. In: "Language" 51, 639-671.

Jelinek, F./Mercer, R. (1985), Probability distribution estimation for sparse data. In: "IBM technical disclosure bulletin" 28, 2591-2594.

Kastovsky, Dieter (1986), The Problem of Productivity in Word Formation. In: "Linguistics" 24, 585-600.

Keune, Karen/Ernestus, Mirjam/Van Hout, Roeland/Baayen, R. Harald (2005), Variation in Dutch: From written MOGELIJK to spoken MOK. In: Corpus linguistics and linguistic theory, in press.

Krott, Andrea/Schreuder, Robert/Baayen, R. Harald (1999), Complex words in complex words. In: "Linguistics" 37, 905-926.

Lüdeling, Anke/Stefan Evert (2005), The emergence of productive non-medical -itis. Corpus Evidence and qualitative analysis. In: Kepser, Stephan & Reis, M. (eds) "Linguistic Evidence. Empirical, Theoretical, and Computational Perspectives". Berlin: Mouton de Gruyter, in press.

Meibauer, Jorg/Guttropf, Anja/Scherer, Carmen (2004), Dynamic aspects of German -er-nominals: a probe into the interrelation of language change and language acquisition. In: "Linguistics" 42, 155-193.

Neuhaus, Hans. J. (1973), Zur Theorie der Produktivität von Wortbildunssystemen. In: Cate, A. P. Cate & Jordens, Peter (eds), "Linguistische Perspektiven. Referate des VII Linguistischen Kolloquiums Nijmegen 1972", Tuebingen: Niemeyer, 305-317.

Nishimoto, Eiji (2003), Measuring and comparing the productivity of Mandarin Chinese suffixes. In: "Computational Linguistics and Chinese Language Processing" 8, 49-76.

Pinker, Steven (1997), Words and Rules: The Ingredients of Language, London: Weidenfeld and Nicolson.

Plag, Ingo/Dalton-Puffer, Cristiane/Baayen, R. Harald (1999), Productivity and register. In: "English Language and Linguistics" 3, 209-228.

Plag, Ingo (2003), Word-formation in English. Cambridge: Cambridge University Press.

Pluymaekers, Mark/Ernestus, Mirjam/Baayen, R. Harald (2005). Frequency and acoustic length: the case of derivational affixes in Dutch. In: Journal of the Acoustical Society of America, to appear.

Pothos, Emmanuel M. (2005), The rules versus similarity distinction. In: Be-

havioral and Brain Sciences 28, 1-49.

Schultink, Henk (1961), Produktiviteit als morfologisch fenomeen. In: "Forum der Letteren" 2, 110-125.

Seidenberg, Mark (1987), Sublexical structures in visual word recognition: Access units or orthographic redundancy. In: M. Coltheart (ed), "Attention and Performance XII". Hove: Lawrence Erlbaum Associates Hove, 245-264.

Skousen, Royal (1989), "Analogical Modeling of Language". Dordrecht: Kluwer.

Wurm, H. Lee/Aycock, Joanna/Baayen, R. Harald (2006), Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. Manuscript submitted for publication.

FIGURE 1

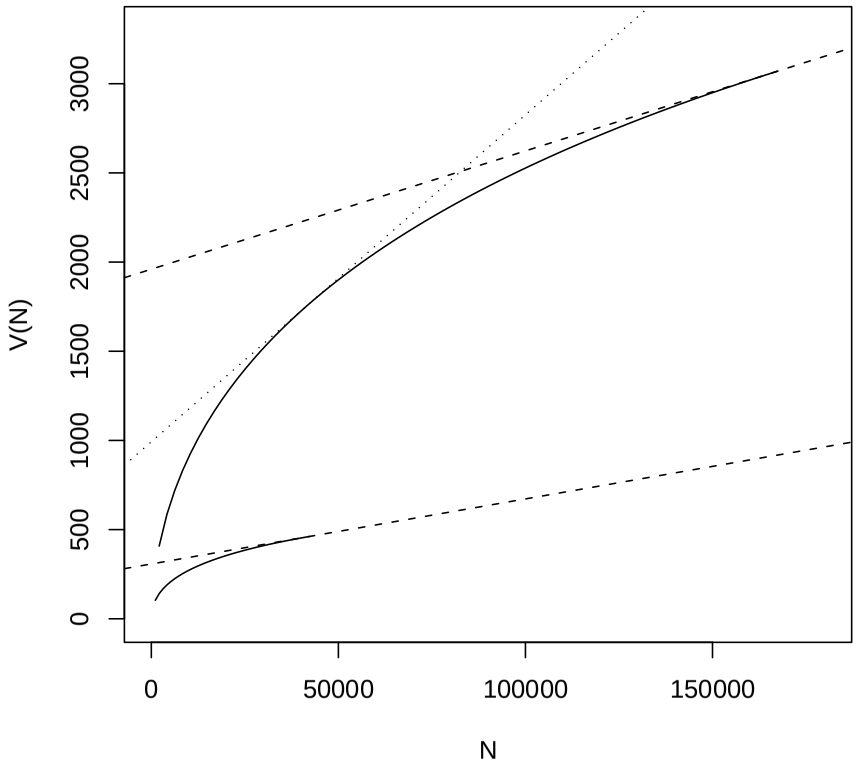


FIGURE 2

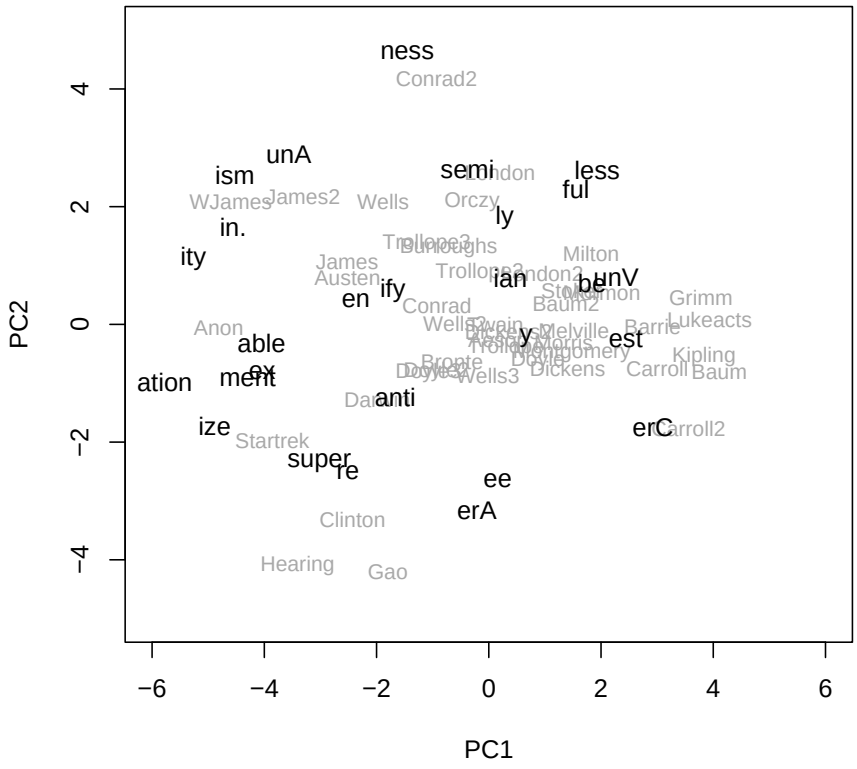

ness
Conrad2
unA
ism
semi
less
ful
WJames
James2
Wells
Orczy
ly
in.
ity
Milton
James
Austen
ify
ian
be
unV
en
Conrad
Baum2
Grimm
Lukeacts
Anon
Barrie
able
Melville
est
Kipling
Carroll
Baum
ation
merit
anti
Wells3
Dickens
ize
erC
Carroll2
Startrek
super
re
ee
erA
Clinton
Hearing
Gao
PC2
PC1
4
2
0
−2
−4
−6
−4
−2
0
2
4
6

FIGURE 3

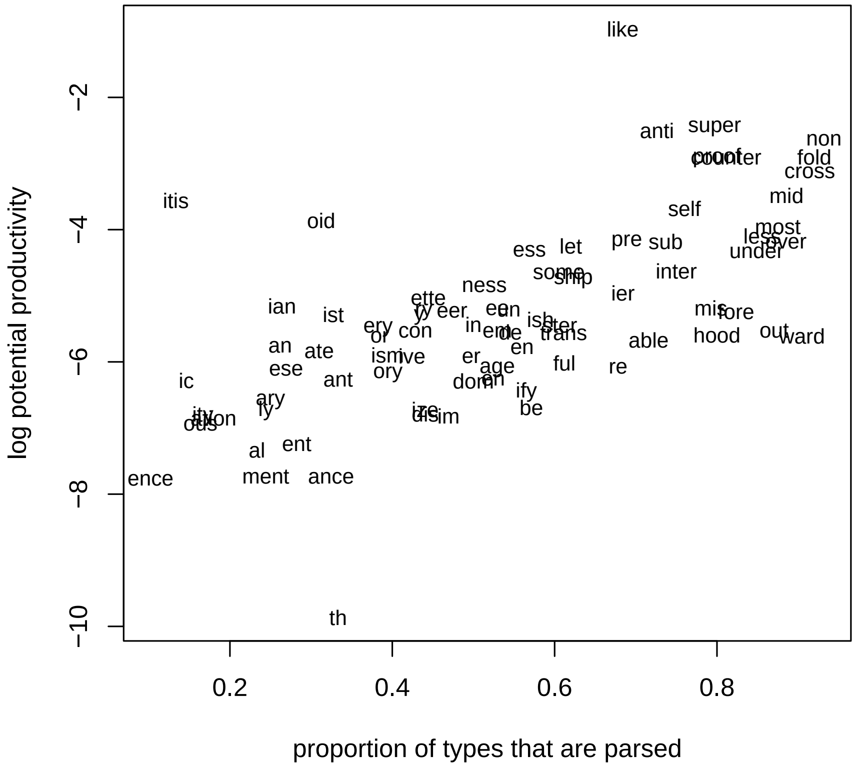
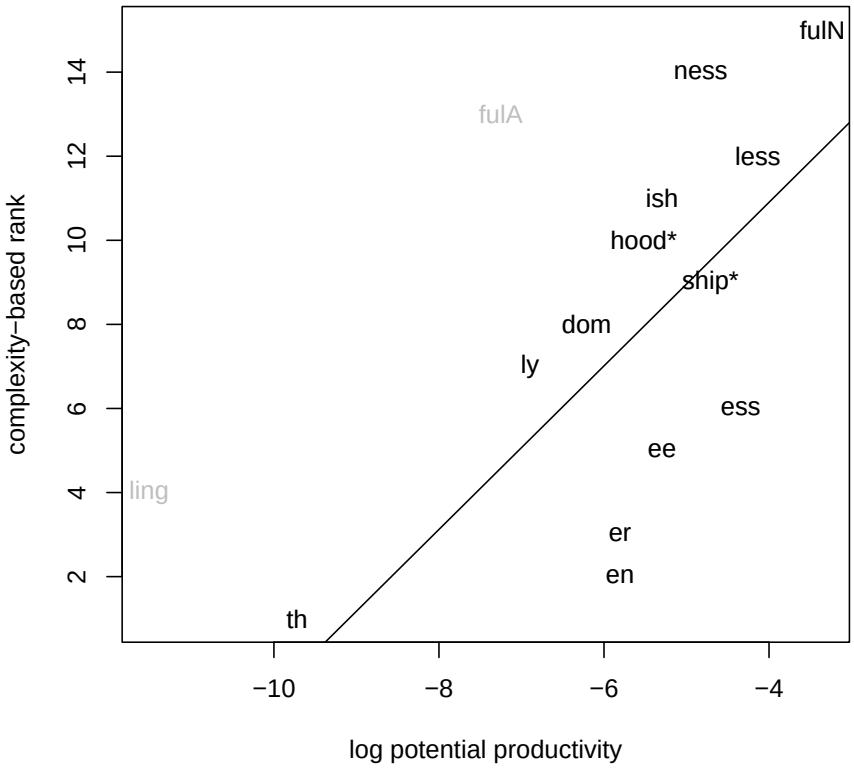

log potential productivity
proportion of types that are parsed
−2
−4
−6
−8
−10
0.2
0.4
0.6
0.8
like
anti
super
non
fold
cross
mid
self
itis
oid
ess
let
pre
sub
most
under
inter
ness
ier
ian
ist
ette
eer
fore
hood
out
ward
able
an
ate
ese
con
ful
re
ic
ant
ory
age
ify
be
ent
al
ence
ment
ance
th

FIGURE 4


fulN
ness
fulA
less
ish
hood*
ship*
dom
ly
ess
ee
ling
er
en
th
complexity−based rank
2
4
6
8
10
12
14
−10
−8
−6
−4
log potential productivity