abstract

This study examines the productivity of five English derivational affixes in the issues of a British newspaper, 'The Times', that appeared in the period September 1989 up to and including December 1992. This diachronic corpus of roughly 80 million word tokens contains large numbers of neologisms. Thus, this corpus offers a good opportunity to test both qualitative and quantitative theories of morphological productivity. Our investigations support the usefulness of the quantitative formalization of the notion 'degree of productivity' developed in Baayen (1992, 1993a). At the same time, they illustrate that productivity is a function of both text type and 'real time'. An investigation of the morphological structure of the neologisms provides strong support for Aronoff's (1976) claim that the productivity of an affix may vary significantly with the morphological structure of the base word to which it attaches.

1 Introduction

The degree of productivity of derivational affixes, the 'statistical readiness' with which these affixes are used to coin novel words (Bolinger, 1948; see also Aronoff, 1980), can be investigated in various ways. Older studies gauge the degree of productivity of affixes by counting the number of words with these affixes in dictionaries (Aronoff, 1976), sometimes supplemented with neologisms collected in texts or colloquial speech (Schultink, 1962). Collections of neologisms in spontaneous speech (Clark & Clark 1979, Clark 1993) are an especially valuable source of information, notably in the domain of language acquisition, where large corpora are as yet unavailable.

Unfortunately, dictionaries are not a reliable source for studying morphological productivity. For instance, Cannon (1988) is a detailed investigation of lexical change in English, based on comprehensive dictionaries such as Merriam-Webster's (1961/1981) Third New International Dictionary of the English Language and its addenda sections. Cannon calls attention to the large numbers of loan words, clippings, blendings, antonomasia, root creation, and Greek and Latin based compounding, compared to rule-governed productive word formation. His counts lead him to conclude that in present-day English the suffix *-ly* is no longer productive: only a handful of neologisms

occur in his dictionary-derived database of neologisms (p. 195).

The conclusion that *-ly* is no longer productive, supposedly having exhausted its potential array of useful possible base words, is unwarranted, however. Dictionaries cannot aim at exhaustiveness in the domain of productive word formation, as it is commercially unattractive to print thousands of words the meaning of which is immediately clear to anyone familiar with the basic meaning of productive affixes. In addition, the compositionality of fully regular morphologically complex neologisms renders their detection by lexicographers especially difficult. Consequently, counts based on dictionaries will often seriously underestimate the extent to which affixes are productive. As we shall see below, the suffix *-ly*, rather than being unproductive, is a very productive derivational affix in present-day English.

Recent developments in corpus linguistics make it possible to study quantitative aspects of the phenomenon of morphological productivity more rigorously and precisely. Probabilistic measures for degrees of productivity, based on the growth rate of the theoretical vocabulary size, i.e., the number of different word types, as a function of the number of word tokens sampled, already tease apart productive and unproductive affixes on the basis of counts in small corpora (Baayen 1992, 1993a). Although experimental elicitation of

neologisms has provided some support for the reliability of these measures (Baayen, 1994b), the large amounts of electronic text that are becoming available at present allow us to compare our probabilistic productivity measures with the rate at which neologisms actually appear over a period of time in written discourse.

In this study, we make use of a database of recent written British English, compiled by the Research and Development Unit for English Studies in the 'AVIATOR' project (see Renouf 1993a, 1993b).[1] AVIATOR had as one of its main objectives the creation of an automated system for the recording of lexical innovation and change. New words, new uses of existing words and the evolving profile of the lexicon were the target areas respectively for a series of software filters. Filter 1 identified the first occurrence of a word in a given database, together with its inception date; Filter 4 amassed a cumulative account of first and subsequent occurrences, with associated frequency counts. With lexicographic applications in mind, the analyzed output was actually held as five, rather than one, lexical inventories; 'ordinary words', proper names, abbreviations, numerals (e.g. '14-year-old'), and queries. 'Ordinary words' are, as expected, deemed to be those consisting of a string of lowercase characters; 'proper names' are those strings with initial capitalization;

and so on.

For the period September 1989 up to July 1993, the frequencies of the words used in 'The Times' newspaper were recorded, as well as their first and last month of use. Exploiting the 'ordinary words' section of this database, we have focused on five de-adjectival derivational affixes, *-ly*, the rival affixes *-ness* and *-ity*, and the rival prefixes *un-* and *in-*, investigating the time course of productive lexical innovation in this newspaper (section 2).[2] In section 3, we focus on the hapax legomena in the database, the relation between hapax legomena and neologisms, and the role of the number of hapax legomena in the mathematical expression for the growth rate of the vocabulary. Structural aspects of lexical innovation are surveyed in section 4. In section 5, we briefly compare the productivity of our five affixes in 'The Times' with their productivity in the 18 million word Cobuild (or Birmingham) Corpus (Renouf, 1987) using the frequency counts of this corpus in the CELEX lexical database (Baayen, Piepenbrock, and Van Rijn, 1993).

2 The time course of lexical innovation

Word use in 'The Times' was registered from September 1989 up to and including June 1993. Our data are based on the first 40 months, the last month taken into account being December 1992. During the first 15 months of the project, only three issues a week were made available in electronic form. In the remaining 25 months, all issues became available.

For each month, the number of types appearing for the first time in that month were registered along with their frequency of occurrence. In addition, the total frequency count of each word was updated for each successive month. The last month in which a word was used, as well as its frequency of use in that month, were also recorded.

PLACE FIGURE 1 APPROXIMATELY HERE

The top panel of Figure 1 plots the number of 'new' words with the suffix *-ly*, word types that appear for the first time in the corpus, for the 40 successive months of our corpus. The observed numbers are represented by dots, the solid line represents a non-parametric regression smoothing (Cleveland,

1979), used to highlight the main pattern in the scatterplot. After the first nine months, hardly any new types occur; it was in these months that fewer issues of 'The Times' were processed. From the sixteenth month onwards, all issues of 'The Times' became available, including all special issues. More than 150 new types in *-ly* alone appear in the sixteenth month of sampling. In the subsequent months, the numbers of new observed types again decrease, as they did for the initial sample of 16 months.

Given that the discontinuity observed at month 16 is caused by sampling factors in the initial phase of the project, we have grouped these initial months into chunks of roughly the size of the monthly chunks obtained from month 16 onwards. The resulting curve of the increase in number of new types as a function of sampling time (in tokens) rather than real time (in months) is shown in the top right hand panel of figure 1. The dots show the observed numbers of new types. The solid line is a non-parametric estimate of the main trend in the data. [3] Observe that the discontinuity in the top left hand graph has disappeared in the top right hand graph. In what follows, we will focus mainly on the distribution of affixes in sampling time, since the irregularities in the numbers of tokens sampled per month are irrelevant for the analysis of productivity.

There are striking differences in the numbers of new types that are sampled after 40 months. After having processed some 80 million word tokens, we encounter 40 new words in *-ly* in the last month of sampling, which amounts to a rate of more than one word each day. Turning to the rival[4] suffixes *-ness* and *-ity* (bottom left panel in figure 1) and to the rival prefixes *un-* and *in-* (bottom right panel), we find that there are hardly any new types for the Latinate affixes *-ity* (11, or 0.37 words per day) and *in-* (0). Their Germanic counterparts appear with substantially higher numbers of new types, 29, or 0.97 words per day, for both *-ness* and *un-*. These rates at which new words are sampled suggest that *-ly* is more productive than *-ness* and *un-*, and that these are in turn more productive than their Latinate rivals, *-ity* and *in-*.

From a psycholinguistic point of view, this is an interesting result. English is a language with very little morphology compared to agglutinating languages such as Turkish or West-Greenlandic. Given the huge capacity of human long-term memory (Landauer, 1986), it seems likely that the storage of complex words is a pervasive phenomenon in the mental lexicon of English language users. Butterworth (1983) and Levelt (1989:185) argue that in everyday spoken English the use of new words is exceptional. This is a strong claim, and there is little doubt that it is correct for normal everyday

conversational English. On the other hand, our data show that some four new de-adjectival words that have not appeared in more than three years of issues of 'The Times' are encountered by readers of this newspaper on a daily basis. Other productive derivational processes, as well as compounding and inflection, may be expected to give rise to even more substantial numbers of additional novel words. As we shall argue in more detail below, words which appear only once in more than three years of newspaper issues are highly likely to be lexical innovations for individual readers of 'The Times'. The non-negligible rate at which novel words appear after processing some 80 million wordforms suggests that productive word formation is not exceptional in written English, which implies that psycholinguistic models of reading should take into account what has been called the 'productivity constraint': such models should not focus exclusively on lexical access to known (possibly complex) words in the mental lexicon, they should also explain how productive neologisms are understood (Frauenfelder and Schreuder, 1992).

Having observed the different rates at which new types appear after screening more than three years of issues of 'The Times', we now turn to consider the lexical statistics of productivity and the relation between 'new types' in our sample and the notion of 'neologism'.

3 Lexical statistics of productivity

The degree of productivity of a word formation rule, which we, following Bolinger (1944), take to be the statistical readiness with which a word formation rule is used to coin or understand new words, can be gauged by means of statistics based on the growth rate of the vocabulary. The vocabulary size V_N, the number of different word types observed after sampling N word tokens of a novel or of a text corpus, is a monotonic, non-decreasing function of N. The rate P at which the vocabulary is increasing for a given sample size N can be estimated using the ratio

$$P_N = \frac{V_N(1)}{N}, \tag{1}$$

where $V_N(1)$ denotes the number of hapax legomena, i.e., the number of types that have been observed only once in the sample. This result from probability theory (Good, 1953) has been used to develop two complementary measures for the degree of productivity of a given affix. One measure focuses on the growth of the full vocabulary, and considers the extent to which different affixes contribute to this overall growth rate. By calculating the proportion $\mathcal{P}^*_{N,c}$ of the number of hapax legomena with affix c ($V_N(1,c)$) on the total of

all hapax legomena h_N among the N tokens of our sample,

$$\mathcal{P}^*_{N,c} = \frac{V_N(1,c)}{h_N}, \tag{2}$$

for different affixes, the likelihood of encountering new words with these affixes after having sampled N word tokens can be gauged.

A complementary measure focuses exclusively on just those N_c word tokens among the N tokens in our sample containing affix c. The growth rate of this restricted sample is given by the ratio

$$\mathcal{P}_{N_c} = \frac{V_N(1,c)}{N_c}. \tag{3}$$

In contrast to $\mathcal{P}^*$, the productivity statistic $\mathcal{P}$ does not take the rate at which tokens with affix c appear in the overall sample into account. What $\mathcal{P}$ represents is the likelihood that, given that a word token with affix c is sampled, this word represents a new type that has not been encountered before. This productivity measure is especially useful for teasing apart affixes that are productive in principle, but that are used relatively seldom, and affixes that are truly unproductive (see Baayen, 1993a).

Both measures stress the importance of large numbers of hapax legomena for an affix to be productive. This ties in with the well-known frequency effect in psycholinguistics. High-frequency words are more likely to be stored

in the mental lexicon than are low-frequency words (Rubenstein and Pollack 1963, Scarborough, Cortese, and Scarborough 1977, Whaley 1978). If a word-formation pattern is unproductive, no rule is available for the perception and production of novel forms. All existing forms will depend on storage in the mental lexicon. Thus, unproductive morphological categories will be characterized by a preponderance of high-frequency types, by low numbers of low-frequency types, and by very few, if any at all, hapax legomena, especially as the size of the sample (corpus or text) increases. Conversely, the availability of a productive word-formation rule for a given affix in the mental lexicon guarantees that even the lowest-frequency complex words with that affix can be produced or understood. Thus large numbers of hapax legomena are a sure sign that an affix is productive.

PLACE TABLE 1 APPROXIMATELY HERE

Table 1 lists the numbers of new types observed in the final month of sampling, as well as the number of hapax legomena in the full sample (the non-normalized $\mathcal{P}^*$ statistic) and the total number of different types V_N. The strong correlation between the number of new types and the number of hapax

legomena is highly significant ($\hat{\rho} = 0.997, t_{(3)} = 23.45, p = 0.0001$, one-tailed test), and higher than the correlation between the number of new types and the total number of types V_N ($\hat{\rho} = 0.907, t_{(3)} = 3.72, p = 0.017$, one-tailed test), confirming the importance of the hapax legomena for obtaining reliable estimates of the rate at which unseen words appear.

The hapax legomena in a sample need not be neologisms, words that are new to the language community and that have not yet been registered in dictionaries and corpus-based word lists. For small samples, nearly all hapax legomena are listed in comprehensive dictionaries. As the size of the sample increases, however, the number of low frequency words that are not listed in dictionaries increases, notably so among the hapax legomena.

PLACE FIGURE 2 APPROXIMATELY HERE

Given the size of our corpus of 'The Times', large numbers of words that are neologisms with respect to a comprehensive dictionary should appear in our sample for the productive affixes. In this study, we have used Merriam-Webster's Third New International Dictionary of the English Language (1961/1981) as a frame of reference for determining whether a word

might be a neologism. Three considerations motivated this choice. First, this dictionary was the most comprehensive synchronic dictionary available to us. Second, Merriam-Webster (1961/1981) is a dictionary that explicitly aims at comprehensiveness (at "maximum coverage with a minimum of compromise", p. 4a). Third, Merriam-Webster (1961/1981) figures prominently in Cannon's (1988) study of lexical innovations, who's claims with respect to productive derivation in current English we disagree with. As will become clear below, even Merriam-Webster (1961/1981) is not an ideal frame of reference for tracing lexical innovation, but at least it is an objective frame of reference for a domain of inquiry where native speaker intuitions fail.[5]

Do we find words in our corpus of 'The Times' that do not appear in Merriam-Webster (1961/1981)? The answer is clearly yes. For *-ness* and *-ity*, 348 and 143 hapax legomena are not listed in Merriam-Webster (1961/1981); for *un-* and *-ity*, we counted 450 and 15 such words, and for *-ly*, 560. These counts are all based exclusively on the hapax legomena, as it is primarily among the hapax legomena that novel words are expected to appear. We have checked this expectation for formations in *-ness* and *-ity*. Figure 2 shows that neologisms, given Merriam-Webster (1961/1981) as frame of reference, typically occur among the lowest type frequencies. For type frequencies 1–5,

figure 2 charts the number of types occurring with that frequency of use in the complete sample. For both *-ness* and *-ity*, the largest numbers of neologisms (grey blocks) appear among the hapax legomena (type frequency 1). Substantially smaller numbers are observed for the dis legomena (frequency 2), and for higher frequencies the numbers of neologisms quickly become negligible. This uneven patterning of neologisms, with increasing proportions of neologisms for decreasing type frequency, with a maximum at the hapax legomena, is highly significant for both *-ness* ($\chi^2_{(4)} = 96.57, p = 0.000$) and *-ity* ($\chi^2_{(4)} = 18.02, p = 0.001$). In other words, the higher numbers of neologisms among the hapax legomena cannot be attributed only to the distributional fact that there are more hapax legomena than dis legomena, and more dis legomena than words occurring three times, etc. Neologisms are significantly overrepresented among the hapax legomena. For larger samples than the one studied here, this tendency for neologisms to appear primarily among the hapax legomena will increase even further. We may conclude that if the productivity of an affix is to be measured in terms of the number of neologisms (defined with respect to a comprehensive dictionary) it gives rise to, then the greatest numbers of neologisms are to be found among the words

with the lowest frequencies of use.

PLACE FIGURE 3 APPROXIMATELY HERE

Up till now we have used a simple criterion to decide whether a word is a neologism. If a word appears in Merriam-Webster (1961/1981), indicating that it has been used elsewhere in the language community, it should be rejected as a genuine lexical innovation. A consequence of this line of reasoning is that words that have been used elsewhere in our sample should also be excluded: any word with a frequency greater than 1 may have occurred in different texts and is likely to represent a word that already belongs to the lexicon of the language community. This leads to a very strict definition of lexical innovation, according to which only those words that are neologisms with respect to existing dictionaries, and that occur once only in large corpora, are counted as lexical innovations.

Figure 3 plots the time course of lexical innovation, defined in this way, for the affixes considered here. The horizontal axis plots sampling time, the vertical axis plots the number of lexical innovations. The solid line is a

non-parametric regression smoother. For the two affixes that are either unproductive or marginally productive, the Latinate affixes *-ity* and *in-*, shown in the right hand panels of figure 3, hardly any lexical innovation is observed. The affixes *-ness, un-* and *-ly*, on the other hand, show much larger degrees of lexical innovation. Especially *un-* and *-ly* reveal regression curves with a positive slope that tends to level off after approximately one third of the tokens have been sampled. In the case of *-ness*, the rate at which lexical innovations appear gradually increases from approximately 8 words per month in autumn 1989 to some 20 words in December 1992, instead of ending up as a constant. This suggests that *-ness* was becoming more productive in these years.

These conclusions are based on a strict definition of productive lexical innovation, a definition that has some disadvantages, however. From a psycholinguistic point of view, the requirement that a word should not be present in a dictionary is unnecessary, as it is highly unlikely that words occurring with a frequency of 1 in 80 million that happen to be mentioned in some dictionary are available in the mental lexicons of individual language users. From a statistical point of view, combining data from dictionaries and a corpus is somewhat unfortunate. Whether or not a regular morphologically

complex word is present in a dictionary is to a large extent arbitrary, there is no common sampling scheme for the dictionary-based and corpus-based counts, and the analysis of the mixed data becomes more complicated, both practically and theoretically.

If we drop the restriction that a word should not occur in a dictionary from our definition of lexical innovation, we are left with a definition that focuses exclusively on the number of hapax legomena with a given affix in the corpus. From a psycholinguistic point of view, this definition is still too strong, in the sense that very low-frequency words that occur more than once in very large samples, such as the dis legomena and the words occurring 3, 4, ..., times in a corpus of 80 million words (see figure 2), are all almost equally unlikely to be listed in the mental lexicons of individual language users — in experimental studies of word recognition, words with a frequency of one per million are already considered to be very low-frequency words. For a discussion of a psychologically motivated productivity measure in which the definition of lexical innovation is based on a broadened range of numbers of low-frequency types and their tokens, the reader is referred to Baayen (1993a).

In the remainder of this paper we will focus on lexical innovation with

reference to the hapax legomena. In addition to the argument that the hapax legomena are used too infrequently to be stored in the mental lexicon, and therefore are prime candidates for rule-based comprehension and production, this choice is motivated by the following considerations. First, the hapax legomena play a crucial role in our probabilistic measures for the degree of productivity. They allow us to infer how many new word types are to be expected for larger samples. Hence their properties are of primary interest from a statistical point of view. Second, as we have seen, most neologisms occur among the hapax legomena. To the extent that dictionaries are a reliable source for determining whether a word is a lexical innovation, an issue to which we shall return shortly, our analyses suggest that the main pattern in the data should already emerge on the basis of the hapax legomena. (Ideally, low-frequency formations that occur just in one text chunk rather than being spread out over different chunks should also be taken into account, as they are also likely candidates for nonce formations. But here the same argument applies: the main pattern is already determined by the number of hapax legomena.) Third, our central concern in this study is with the spontaneous, unintentional and ephemeral use of productive word formation, not with conscious and deliberate lexical creativity in which a novel expression is carefully

constructed to express a new concept intended for repeated use within an — often specialist — topic domain. The hapax legomena in our corpus tend to be the prototypical instantiations of ephemeral word formation that is most similar to the productivity of syntactic constructions.[6]

Looking at this 'syntactic' aspect of morphological productivity, it is useful to consider how the words that are hapax legomena in the complete sample (the 'global' hapax legomena, as opposed to the 'local' hapax legomena $V_M(1), M < N$) are distributed in sampling time. Figure 4 plots the number of new types sampled (solid lines), and the number of types that are global hapax legomena, i.e., that occur only once in the full sample of 80 million tokens (dotted lines), as a function of the number of tokens sampled. First note that the two curves converge: in the final month of sampling, nearly all new types are words that are used once only. From a probabilistic point of view this is to be expected. From a linguistic point of view, however, it is non-trivial, as it clearly demonstrates that new words can be used ephemerally, without repetition and further elaboration. Although neologisms can be created to fill a lexical gap, as when new technical terms are introduced to describe novel concepts, in which case they are likely to be used more than once, this appears to be a rare phenomenon in our corpus. What our

data suggest is that the productively coined lexical innovations that one will observe after having sampled millions of word tokens are nonce formations that exemplify the combinatorial possibilities of the language in the same way as syntactic structures.

PLACE FIGURE 4 APPROXIMATELY HERE

More importantly, figure 4 suggests that for *un-*, *-ness* and *-ly*, but not for *in-* and *-ity*, the number of global hapax legomena increases as a function of sampling time (in tokens). This is shown more clearly in figure 5, which presents the individual numbers of global hapax legomena sampled, a non-parametric regression curve (solid line), and the least squares regression line (dotted line). Under random conditions, one would expect the hapax legomena of the full text to be randomly (i.e., uniformly) distributed over sampling time. To check this for a large single text, we analyzed the distribution of hapax legomena in H. Melville's *Moby Dick* ($N = 213756$, $V_N = 16741$, using a wordform-based definition of word type). The bottom left-hand panel shows the pattern of results obtained. The scatterplot suggests a random

pattern, with the exception of the last text chunks, where hapax legomena appear to be underrepresented.

The parametric least squares analyses for the Germanic affixes all show significantly positive correlations and slopes ($F_{(1,25)} = 5.087, p = 0.011$ for *un-*, $F_{(1,25)} = 29.42, p = 0.000$ for *-ness*, and $F_{(1,25)} = 15.54, p = 0.001$ for *-ly*), whereas there is no correlation between number of hapax legomena and sampling time for *in-* and *-ity* ($F_{(1,25)} = 1.19, p = 0.24$ for *in-*, $F_{(1,25)} = 2.25, p = 0.146$ for *-ity*). The negative slope of the regression line for *Moby Dick* is not significant ($F_{(1,38)} = 1.545, p = 0.221$). The positive slopes for the regression lines obtained for the productive affixes suggest that these affixes may be becoming increasingly productive in 'The Times'. This would imply that the urn model for word frequency distributions, according to which word tokens are randomly sampled from a fixed population of types with known and fixed population probabilities (see Baayen 1993b, 1994c, Chitashvili and Baayen, 1993) is incorrect. What our data may suggest is a dynamic system in which the population probabilities change over time. If correct, the productivity measures discussed above are approximations rather than true estimates of the growth rate.

Figures 3 and 5 suggests that *-ness*, *un-* and *-ly* are becoming more pro-

ductive in British English. The idea that we are observing diachronic change in productivity has to be handled with caution, however, as it is possible that the non-uniform distributions of hapax legomena are the result of changes in the population of journalists writing for 'The Times'. Changes in the number of (freelance) writers, changes in the number of events that evoke or inspire lexical innovation, or changes in the writing skill of these journalists might be involved. The fact that the non-parametric regression lines level off by the end of the sample suggests that perhaps we are observing a ceiling effect in writing skill. Clearly, the observed non-random pattern in the distribution of hapax legomena in *-ness*, *-ly*, and *un-* needs to be replicated in other diachronic corpora before any conclusions with respect to changes in productivity in English can be drawn with confidence.

PLACE FIGURE 5 APPROXIMATELY HERE

Our finding that *-ly*, *-ness*, and *un-* are clearly productive contrasts sharply with the conclusions reached by Cannon (1988) on the basis of a survey of neologisms in dictionaries such as Merriam-Webster (1961/1981). Whereas our investigation clearly shows that the Germanic affixes enjoy a

much higher degree of productivity than the Latinate affixes, Cannon's counts lead him to conclude that in English 'Most productive suffixes now come from Latin or French' (p. 195). For instance, Cannon's counts reveal only 5 neologisms with *-ly*, and more neologisms with *-ity* (28) than with *-ness* (10) (p. 194–195). The only possible explanation for this mismatch between dictionary-based and corpus-based data is that, apparently, lexicographers are more likely to observe and list new words when these words contain less productive or unproductive affixes.

This ties in with a defining criterion of productive word formation, unintentionality of use (Schultink, 1962). In the same vein, Aronoff (1982) points out that the more productive a word formation pattern is, the less likely a speaker or listener will be able to distinguish between new formations and existing ones. Conversely, as the productivity of a pattern decreases, the likelihood increases that speakers are aware of the fact that they are coining a new word. They may even exploit the salience of semi-productive neologisms for foregrounding purposes.

From this perspective, the tendency for lexicographers to list words with less productive affixes more often than words with fully productive affixes is only to be expected. A high degree of productivity makes it difficult to

recognize a neologism as such, and this decreases its likelihood of being represented in a dictionary. Even dictionaries which aim at comprehensiveness, such as Merriam-Webster (1961/1981), cannot reach their objective as long as neologisms have to be gathered by means of human observation. Only computerized processing of large-scale corpora will allow us to obtain a complete overview of the quantitative aspects of lexical innovation.

Cannon also suggests that normal derivational processes give rise to relatively few neologisms in comparison to word formation involving creative processes such as blending, clipping, and learned Latin and Greek-based compounding. In his database, only 24% of all neologisms are the result of regular affixation (p. 191). Although we have only begun to explore English word formation in 'The Times', our data strongly suggest that Cannon's dictionary-based study underestimates the extent to which rule-governed derivation gives rise to neologisms. Whereas Cannon (1988:194) observed 10 neologisms in *-ness* on the basis of dictionary-based research, our database contains more than 350 formations in *-ness* that are not listed in the Webster's Third New International Dictionary of the English Language (1961), nor in Merriam's 1981 addenda section to this dictionary.

4 Structural aspects of lexical innovation

Having discussed the quantitative aspects of the time course of lexical innovation in some detail, we now discuss the structural aspects of lexical innovation, focusing on the way in which the productivity of affixes varies with the morphological constituency of the base word.

4.1 -ly

Our database for all 46 sampling months of the AVIATOR project contains 5196 word types with the suffix *-ly*, 1362 of which are hapax legomena. Concentrating on the hapax legomena as prime instantiations of lexical innovations, we find 82 instances of adjectivizing *-ly*, and 1280 instances where *-ly* forms adverbs. Clearly, adjectivizing *-ly* is much less productive than adverbializing *-ly*.

This difference in productivity becomes even more apparent when we consider the kind of base words to which *-ly* attaches. The nouns to which *-ly* is suffixed to form adjectives can be divided into compounds (11 cases, e.g., *headmistressly*), monomorphemic nouns (64, e.g., *lizardly*), prefixed nouns (1, *archbishoply*), and suffixed nouns (6 with *-er*, *preacherly, loverly, shutterly, composerly, slipperly*, and *reporterly*, and 2 with *-ist*, *defeatistly, escapistly*).

Apparently, denominal *-ly* is hardly productive when attached to derived nouns — among the words occurring more than once, only 6 other base words end in *-er* (*designer, reader, dancer, teacher, writer, painter*), and none in *-ist*. This suggests that the tendency for denominal *-ly* to attach mainly to concrete nouns (Quirk, Greenbaum, Leech, and Svartvik, 1985, p. 1002), is so strong that not only abstract nouns such as *speed* and *peace*, but also non-institutionalized agent nouns such as *thinker* and *jumper* are non-optimal as base words.

Adverbial *-ly*, on the contrary, is extremely productive with respect to a wide range of base word types, both simplex (175, e.g., *blondely*, a category in which we have also included words with possible bound stems, such as *episcopally* and *tumescently*), and complex (e.g., *-al*: 213, *-ing*: 183, *un-*: 157, *-ed*: 74). There are 30 synthetic compounds with *-ing* (*breathcatch-ingly*), 4 instances where *-ly* attaches to a phrase (*themisvously*, *real-worldly, out-of-breathly, unchanged-to-modestly*), and 4 cases of affix generalization (*abroadly, whyly, onely, oftenly*). The top right-hand panel of figure 6 summarizes the way in which the productivity of *-ly* varies with the structure of its base words.

There is one kind of base word that does not allow adverbial *-ly* to attach,

namely adjectives formed by means of *-ly* itself. There has been some discussion concerning the nature of this restriction (Aronoff, 1976:37fn; Quirk et al., 1985; Bauer, 1992). Quirk et al. (1985) state this restriction as a categorical rule, although they acknowledge that 'occasionally, one finds *sillily, friendlily*' (1985:1556). Bauer (1992) calls attention to a small number of such formations in the OED (1987), arguing that even when absolute restrictions on productivity are in force, productivity remains a scalar notion. In the case at hand, even though adjectives in *-ly* are ruled out as potential base words for adverbial *-ly*, Bauer argues that *-ly* retains some low degree of productivity here.

In our corpus, the only two words ending in the string *lily* are *friendlily* and *lovelily*. Both occur only once. The absence of any other *lily* adverbs in the higher frequency ranges suggests that we are dealing with ephemeral formations that are in no way supported by words with a similar structure that are well-entrenched items of the English lexicon, let alone by a word formation rule. Theoretically, their status is similar to the instances of affix generalization mentioned above, where *-ly* is similarly applied to base words that fall outside its normal input domain. In other words, *friendlily* and *lovelily* are exceptions to a rule rather than the result of a rule with a very

low degree of productivity. The idea that the notion 'degree of productivity' comes into play only for the — categorically defined — input domains of a word formation rule (Schultink, 1962) cannot be rejected on the basis of adverbial *-ly*.

This illustrates an important methodological point, namely that the hapax legomena should not be studied in isolation from the rest of the frequency spectrum. Conditional on the set of all *lily* adverbs in our database, the degree of productivity as measured by $\mathcal{P}$ would equal unity, suggesting full productivity for this input domain of *-ly*. If an input domain **only** yields very low frequency words, however, this may indicate either overextension of a rule, or the loss of a rule, in which case the low-frequency words are the last remnants of that rule that are in the process of disappearing from the language.

4.2 -ness

The suffix *-ness* is instantiated in 2027 different words in our database, 739 of which are hapax legomena. Among the latter, the largest number of lexical innovations (138) is derived from monomorphemic basewords (*floridness, leftness, pinkness*), a category in which we have also included base words

with an affix attached to (synchronically opaque) bound stems (*glottalness, emotiveness, literateness*). Some formations illustrate that blocking is not an absolute phenomenon (see Rainer, 1988): *curiousness* occurs once in spite of *curiosity* being the default choice, and the same holds for *secretness* and *secrecy*, and *vivaciousness* and *vivacity*. Such blocking failures are quite common. While at first sight these formations appear to be redundant, since an established alternative exists, they are often created to mean something slightly different.

Adjectives in *-y* also give rise to large numbers of lexical innovations (130). Some of these have the simple structure N+y+ness, such as *crabbiness, nerdiness, twittiness, zappiness, iffiness*, others involve a phrase which is transposed into an adjective by *+y*, after which *-ness* is attached: *chalk-and-cheesiness, outdoorsiness, show-offiness*. Compound N+A adjectives such as *fashion-conscious, streetwise*, and *tone-deaf* (34), and A+N+ed adjectives such as *sticky-fingered, surehanded* and *male-centered* (56) are also well-represented.

Other adjectivizing affixes such as *-ish* (*babyish*), *-less* (*shoelessness*), *un-* (*unbedworthyness*), *over* (*overanxiousness*), and *-ed* (*datedness*) are used less extensively (*-ish*: 27, *-less*: 37, *un-*: 40, *-ed*: 30, *over-*: 15). Interestingly, of

the 15 formations with *over-* in our data, 14 are not listed in Merriam-Webster (1961/1981), suggesting that it is becoming more productive, perhaps along with vogue prefixes such as *ultra-* (*ultra-selfishness*), *hyper* (*hyper-alertness*) and *super-* (*super-alertness, super-coolness, super-richness*).

Of the affixes studied here, *-ness* shows the highest degree of affix generalization (24, e.g., *cowness, duckness, endness, ladyness, godness, redneckness, wonderness, footballness, joyness, itness, on-ness*, *ex-ness*). Although affix generalization is not unattested in Merriam-Webster (1961/1981) (*outsideness, togetherness*), it occurs more often than a dictionary-based survey would suggest, perhaps because the phenomenon has been frowned upon in the authoritarian tradition of lexicography (and criticized by Williams (1965), who reports the use of denominal formations such as *carness, dayness* and *verbness* and many other instances of affix generalization such as *hereness, inness, thereness* and possibly *wellness*). We also find examples of *-ness* attaching to phrases (16, e.g., *next-to-nothingness, thatitness, over-the-top-ness, olde-worlde-ness*).

When we compare the numbers of instances of affix generalization observed for *-ly* and *-ness* within their respective sets of hapax legomena, we find that affix generalization occurs more often with *-ness* (24/758) than

with *-ly* (4/1362), even though the latter is more productive than the former ($\chi^2_{(1)} = 28.67, p = 0$). This suggests that it is the semantics of an affix rather than a high degree of productivity as such that drives affix generalization.

In comparison to *-ly*, *-ness* is only weakly attested for base words in *-ing* (13, e.g., *demandingness*) and synthetic compounds with *-ing* (5, e.g., *eye-catchingness*), in all 18 out of 758 hapax legomena. By contrast, we have counted 183 adverbs in *-ingly* and 30 cases where *-ly* is suffixed to a synthetic compounds, 213 out of 1362 hapax legomena. The difference between the two proportions is highly significant ($\chi^2_{(1)} = 86.88, p = 0.00$). This suggests that the moment of abstraction in the semantics of *-ness* does not lend itself well for combination with the event or change-of-state semantics of formations in *-ing*.

4.3 -ity

The suffix *-ity* appears with 1020 different word types in our database, of which 280 are hapax legomena. The highest degrees of productivity are observed for base words in *-able*, (77) followed by monomorphemic base words (50, *anality, loyality, concavity*), a category in which we have also included base words with bound stems (*avuncularity, spectacularity, deviosity*). Other

base types that are reasonably well represented are *un-* (28), *in-* (18), and *-al* (23). As in the case of *-ness*, we observe a fair number of affix generalizations (12, e.g., *assurity, terrority*). The proportions of affix generalizations among the hapax legomena do not differ significantly for *-ness* and *-ity* ($\chi^2_{(1)} = 0.37, p = 0.54$), confirming the hypothesis advanced above that it is the semantics of a word formation process that will determine the amount to which affix generalization will take place, and not its degree of productivity as such.[7]

4.4 un-

The prefix *un-* occurs in 1672 different word types, 659 of which are hapax legomena. Of these 659 hapax legomena, 3 formations are denominal (*un-conclusion, unattraction, undabber*), 10 are deverbal (*unsay, unbuild, undock, unsling, unseam, unbunch, unrighten, unspread, unmuck, uncaress*, of which the latter five are not listed in Merriam-Webster, 1961/1981), and 646 are de-adjectival. The highest degree of productivity is found for base words in *-ed* (294, e.g., *unnannied*), followed by monomorphemic base words (88, e.g., *unflat, ungreen, unserene*), *-able* (77, e.g., *uncorkable*), *-ing* (57, e.g., *unshrinking*), and *-y* (29, e.g., *uncheesy, unsuperstarry, unfishy*). There is

only a single instance of affix generalization (*unabove*), and one formation that might be analyzed as involving a phrasal base (*unsetup*).

These counts suggest that *un-* hardly allows affix generalization. Even if we include the denominal formations, only 4 instances out of 648 are counted. Note that a single base type (*X-ed*) accounts for almost half of all hapax legomena. This shows that the variation in degree of productivity across different input domains can be very large indeed.

In addition, the small number of deverbal hapax legomena highlights that the productivity of deverbal *un-* is of another order of magnitude compared to that of de-adjectival *un-*. Although deverbal *un-* is not totally unproductive, very few additional formations are to be expected in additional issues of 'The Times'. Its degree of productivity is so low that it is not clear whether it can be spontaneously attached to new base words. For instance, Levin (1993:167) mentions that other members of the class of *tape* verbs 'might allow' for prefixation with 'reversative' *un-*, which implies that it is not self-evident that all members of this class are unambiguously available as base words. Interestingly, our database contains two formations, *unbelt* and *unstrap*, illustrating that Levin's class of 'disassemble' verbs can indeed be extended. At the same time, corpus-based inference and native speaker

intuitions reveal that extension is very unlikely to occur.

Why is verbalizing *un-* semi-productive? First, verb formation in general may be a more complex process than noun formation with *-ness* and *-ity* or adjective or adverb formation with *-ly*, as it involves the setting up of, or the modification of, predicate-argument structure. This may render the coining of new complex verbs more difficult, leading to lower degrees of productivity (Baayen and Lieber 1991:836). Second, in the case of verb-forming *un-*, many potential reversative meanings are already expressed by other verbs. For instance, *unclose* is blocked by *open*, *uncome* by *go*, *unattach* by *detach*. The prefixes *dis-* and *de-* in *deconstruct* and *dismantle* block prefixation of *un-* to words such as *construct*. Third, many actions are irreversible in everyday context of use. For instance, murder cannot be undone, except in time-travel science fiction, where one can *unmurder, unmarry* and even *uncreate* (Kastovsky 1986:594). This leaves *un-* free to attach to a fairly limited class of primarily verbs of attachment and covering, verbs that express actions that are easily conceivable as reversible.

Bowerman (1982) and Clark (1993), following Whorf (1956), argue that when *un-* is attached to verbs to coin verbs, its use is governed by a condition which restricts its input to verbs of attachment and covering. This

covert condition, Bowerman argues, poses a learning problem for children. Her data suggest that, after having mastered a few central deverbal *un-* formations without analyzing them into their constituents, children may learn the reversal semantics of *un-* before they master the covert condition restricting its use to verbs of attachment and covering. This, she argues, is revealed by overextensions such as *unstraighten* and *unhate*, where *un-* is attached to verbs falling outside the adult input domain.

Bowerman's explanation raises the question as to why adults may be uncertain about the appropriateness of prefixing *un-* to a new attachment or covering verb. Once a restriction on a word formation rule has been mastered, uncertainty should be absent, and new formations should be freely coinable. This does not seem to be the case, however. Why not?

In our view, there is no such thing as a restriction limiting *un-* to attachment and covering verbs. Although this class is the most natural source of base words, other kinds of base words are in use, such as *(un)do*, *(un)learn*, *(un)make*, *(un)freeze* and *(un)say*. Occasional neologisms such as *unmarry, unmurder* and *uncreate* are quite natural in settings in which their meanings make sense, and our database contains two hapax legomena that are not listed in Merriam-Webster (1961/1981) that fall outside the attachment

and covering class, *unrighten* and *uncaress*. Interestingly, Bowerman's (1982) examples of children's overextensions fall into two classes, blocking failures (*unstraighten* instead of *bend*, compare *unrighten* in our database), and formations expressing not reversal but the end of a process (*unfuzz* and *unhate* are intended to mean 'stop fuzzing', and 'stop hating' respectively). The latter type of overextension suggests that the reversative meaning of *un-* has not yet been mastered by the child who produced these words.[8] These formations do not provide genuine examples of overextensions of reversative *un-*, and the same holds for the blocking failures. This suggests that no specific condition restricting *un-* to verbs of attachment and covering is required to explain its low degree of productivity — the scarcity of base words expressing a naturally reversible action for which no established word is already in current use is responsible.

4.5 in-

Our database contains 243 different words with the prefix *in-*, 48 of which are hapax legomena.[9] Among these hapax legomena, we find formations where the rival form with *un-* seems to be more natural (e.g., *inuseful, incertain, inequal, injustified, insteady, indisputed*), all of which suggest failure of block-

ing by *un-*.

Other hapax legomena appear to be well-established technical terms that happen not to enjoy a high frequency in 'The Times' (e.g., *incompetence, independence, inconvenience, incongruent, indistinctive, inestate*). Examples of genuine lexical innovations in *in-* are *ineliminable, inexpiable, inconvertable, inegalitarian, inignominious.*

The small overall number of hapax legomena, and the fact that less than half of these involve genuine lexical innovations, shows that *in-* is hardly productive in the newspaper English of 'The Times'.

4.6 Discussion

We have briefly described the structure of lexical innovations for 5 affixes, all of which predominantly attach to adjectives. We have seen that the number of hapax legomena may vary substantially with the structure of the base words to which affixes attach, and we have implicitly assumed that these variations can be understood as differences in the degree of productivity of an affix across its input domains.

The hypothesis that productivity may vary with the morphologically defined input domains of a word formation rule is attributable to Aronoff (1976)

and, independently, to Kjellmer (1984). It has been seriously criticized, however, by Van Marle (1985). Van Marle is an exponent of the European structuralist tradition, in which degrees of productivity are claimed to reflect the extent to which phonological, morphological, and semantic restrictions constrain the input domain of an affix. For instance, the restriction that bars *-ity* from attaching to base words from the so-called native stratum of the lexicon renders *-ity* less productive than *-ness*, which is free to attach to both native and non-native base words. Van Marle (1985) argues that as the number of such restrictions increases, the degree of productivity decreases. Within the input domain defined by these restrictions, an affix is claimed to be absolutely productive. Any numerical differences within this input domain, such as widely varying numbers of types and neologisms as a function of the morphological structure of the base word, are argued to reflect differences in what Van Marle calls degrees of actuation. In other words, the restrictions defining the input domain of an affix are viewed as originating from lexical competence, and define the productivity of a rule. Any remaining quantitative variation is declared to be a matter of performance, which allows different degrees of actuation.

In our view, this approach to morphological productivity is fundamen-

tally flawed, for a number of reasons. First, the idea that the degree of productivity decreases as the number of phonological, morphological, and semantic restrictions on a word formation rule increases is too imprecise to have any quantitative validity, for the simple reason that the number of different types removed from the input domain of an affix by such restrictions may vary widely from restriction to restriction. Without additional qualification of the restrictive weight of these restrictions, the claim that the degree of productivity and the number of restrictions are inversely related is simply vacuous.

Consider, for instance, the restriction that prohibits adverbial *-ly* from attaching to adjectives in *-ly*, ruling out a number of what would otherwise have been potential adverbs in *-ly*. We do not know how many *-lily* adverbs might have appeared in our corpus if no restriction had been in force. Given the small numbers of derivations with *un-* and *-ness* from adjectives in *-ly*, however, it is rather unlikely that this restriction rules out large numbers of complex words. The so-called 'difference in degree of actuation' between words of the form *over-X* and *X-ing* for adverbial *-ly* probably have much more quantitative weight.

Second, there are substantial differences in the numbers of types as a

function of the morphological structure of the base word for all the affixes studied here, both across affixes and within the input subdomains of each of the affixes separately. Substantial variations can be observed for the base types of a single affix, as is immediately apparent from figure 6. For instance, base words in *-ed* are a very productive source for words in *un-*; adjectives in *-ly*, on the other hand, are seldom preceded by this prefix.

PLACE FIGURE 6 APPROXIMATELY HERE

Similar differences can be observed for a fixed kind of base word across the five affixes. For instance, the number of derived hapax legomena for base words in *X-al* varies from 0 for *in-* to 213 for *-ly*. Figure 7 summarizes the differences in productivity for affixes and base types jointly by means of a chi-plot of a chi-square test of independence. The position of a bar, above or under the line, indicates whether there are more or fewer items than expected given the marginal numbers of hapax legomena. The size of a bar is proportional to the extent to which a particular base word is overrepresented or underrepresented. More precisely, the (squared) height of a bar is proportional to the contribution to the chi-squared statistic, the

width of a bar is proportional to the square root of the expected number of hapax legomena, and the area of each bar is proportional to the residual deviance, the difference between the observed and expected number of hapax legomena.

Figure 7 is especially revealing in that it shows that the quantitative variations within and across affixes as a function of the kind of base word is not simply determined by the numbers of available base words and the overal productivity of the affixes.[10] It is easy to see that base words in *-al*, *-ing*, and *un-* are especially productive for *-ly*, but that the large number of monomorphemic words to which *-ly* attaches (see figure 6) is to be expected given the high overall productivity of *-ly* and the numbers of monomorphemic base words to which the other affixes attach. Base words in *-y* are most likely to give rise to formations in *-ness*, but *-ness* is not particularly productive for base words in *-ed, -able, -al*, and *-ing*.[11] Note that *-ness* and *-ity* are the only affixes that appear with positive scores for lexical innovations with phrasal bases and base words containing the prefix *over-*, and that only these affixes reveal positive bars for affix generalization.

These often large and highly significant quantitative differences exemplified in figure 7 need to be explained, even though it is not always immediately

obvious what forces might give rise to these differences.[12] To claim that they are due to differences in degrees of actuation amounts to sweeping the phenomenon at hand under the carpet as a 'performance' issue irrelevant to the study of language structure.

PLACE FIGURE 7 APPROXIMATELY HERE

This brings us to our third, and most important point, namely that the structuralist approach to morphological productivity fails to take into account that word formation is conceptually driven, and that the restrictions defining a word formation rule only set the boundary conditions for word formation. Many such restrictions rule out phonotactically illegal or sub-optimal sequences that might arise through affixation (the restriction barring adverbial *-ly* from adjectives in *-ly* is a possible example). They similarly prevent the formation of words that are semantically or pragmatically infelicitous. It is essential to realize that these restrictions impose limits on word formation processes that are **conceptually** driven. The function of word formation is to convey (particular shades of) meaning, not simply to produce forms with a particular structure. A description of the domain of possible input

words amounts to a definition of a word formation rule (Van Santen, 1992), the point of departure for an analysis of its productivity. Depending on the semantic structure of the base word, of which the morphological structure is a crucial determinant, the semantic operation expressed by a particular affix may be more or less natural and useful. From this perspective, it is only to be expected that the effects of the boundary conditions are outweighed by the conceptual forces driving word formation.

5 Comparison with Cobuild/CELEX database

Hitherto, our discussion has been based on a fairly restricted kind of English, written English in one particular newpaper, 'The Times'. However, the way in which words are put to use may vary substantially from genre to genre, from text type to text type, and even from author to author (see, e.g., Burrows, 1992; Biber, 1989; Biber and Finegan, 1989). This raises the question to what extent our results would generalize to other kinds of English, the more so as there is some evidence that suggests that the productivity of affixes is similarly subject to variation as a function of text type and style (Baayen, 1994a). As a first approximation, we have compared the

frequency distributions obtained for 'The Times' with the corresponding frequency distributions based on the Cobuild Corpus (Renouf, 1987) that, at the time that its frequency counts were made available to the CELEX lexical database, contained some 18 million word tokens, of which 1.3 million were spoken English word tokens. The texts underlying the Cobuild/CELEX counts represent a much broader variety of text types, and this may give rise to different productivity profiles for our affixes.

The differences in the numbers of tokens sampled make it impossible to compare directly the frequency distribution of, e.g., *-ness* in 'The Times' with the corresponding frequency distribution of *-ness* in the Cobuild/CELEX database. However, the two frequency distributions can be used to interpolate the number of different types V_N for smaller sample sizes. Figure 8 plots the growth curves V_N for *-ness, -ity, un-, in-* and *-ly*. The solid lines represent the interpolated values of V_N as calculated on the basis of the frequency distributions in 'The Times'. The dotted (shorter) lines have been obtained on the basis of the Cobuild/CELEX counts. Where the two curves are more or less superimposed, we do not have any reason to suppose that differences in productivity are at stake. If the two curves differ substantially, the one with the higher slope represents the corpus in which the affix is more

productive.

Figure 8 shows that the curves for *-ness* and *-ity* are very similar. This indicates that their productivity does not differ substantially in 'The Times' and in the Cobuild/CELEX database. In the case of formations in *-ly*, the Cobuild Corpus reveals slightly larger numbers of different types. For the prefixes *un-* and *in-*, however, the two curves diverge substantially. The Cobuild/CELEX sample yields roughly the same number of different types in *un-* as does 'The Times', even though the latter sample is approximately four times the size of the former. A similar situation obtains in the case of *in-*. Clearly, negative prefixation is much more productive in the Cobuild Corpus than in 'The Times'. The inclusion of a greater variety of text types in the Cobuild Corpus (including literary texts) is likely to be the main source of this difference. If so, it is remarkable that predominantly negative prefixation is affected, and possibly suffixation of *-ly* to some extent, but not de-adjectival nominalization.

Note that both *un-* **and** *in-* reveal this marked difference in productivity. This provides additional support for our claim that the semantic function of affixation, negation in the case of *un-* and *in-*, is a strong determinant of

productivity.[13]

PLACE FIGURE 8 APPROXIMATELY HERE

6 Discussion

According to Clark (1993), analyses of morphological productivity should proceed on the basis of lexical innovations only. Established, conventional words that are already in the lexicon, she argues, 'are liable to display all the idiosyncrasies of forms and meanings that accrue over time' (p. 128). The present study supports Clark's claim that analyses of productivity based on counts of actual words listed in lexicons are unreliable. For instance, a comparison of Cannon's dictionary-based study with the present corpus-based approach reveals that word formation in the native stratum of the lexicon is much more productive than dictionaries would suggest.

The analyses presented here are firmly based on a survey of neologisms and as such meet Clark's strict criterion for the study of productivity. Nevertheless, we feel that Clark's criterion is too restrictive, for two reasons. First,

statistical inference on the basis of relatively small corpora already allows the main trends in productivity to emerge on the basis of actual words in the lexicon. This claim, advanced in Baayen (1993a), is supported by the present analyses of our corpus of 'The Times', a corpus that is large enough to reveal extensive use of neologisms. Second, idiosyncrasies of meaning are not restricted to existing words. The intentional foregrounding often underlying the use of semi-productive word formation may give rise to lexical innovations that are not fully regular. For instance, in 'A woman partial to clothes that outlined, profiled, unimagined the body.' (B. Pesetsky, The Late Night Muse, Harper Collins, New York, 1991, brought to our attention by Beth Levin), *un-* in *unimagined* illustrates intentional, creative literary use of a semi-productive word formation pattern. In this example, *un-* does not denote the reversal of the action of imagining, but that imagination is rendered unnecessary. Both actually existing words and lexical innovations require careful study of their meanings if the semantics of productive and semi-productive word formation rules are to be understood.

By 'chronicling' The Times for a number of years, we have shown that even after sampling some 80 million wordforms, substantial numbers of de-adjectival lexical innovations in *-ly, -ness, un-* and to some extent *-ity* ap-

pear. Most lexical innovations are nonce formations enjoying ephemeral use. Where fully productive word formation is concerned, the difference between morphological and syntactic productivity is a quantitative, not a qualitative one.

Whereas adjectivizing *un-* was found to be highly productive, verb-forming *un-* turned out to be semi-productive. Clearly, phonological and semantic transparency is a necessary, but not a sufficient, condition for productivity. In the case of verb-forming *un-*, the highly limited number of verbs expressing reversible actions for which no reversative predicate is already available in the language is probably one of the reasons that verb-forming *un-* is not fully productive. Another reason is that the creation of reversative verbs involves operations on argument structure.

Interestingly, word formation can be exploited for purposes that go beyond concept formation. Suffixation of *-ly* primarily serves the syntactic purpose of explicitly marking adjectival modification of a predicate instead of a noun. Concept formation is much more clearly involved in suffixation with *-ness*. Nevertheless, *-ness* often serves the purpose of event reference, as in *He was very indecisive. In fact, his indecisiveness led them to ...* (see Kastovsky, 1986). De-adjectival *un-* may be used similarly in discourse, but

in addition, prefixation with *un-* cannot be exchanged for syntactic negation with *not* without changing the implicature of the sentence (Horn, 1984). This implies that *un-* is also functional on the level of pragmatics. Given that the use of adverbial *-ly*, de-adjectival *un-*, and *-ness* is co-determined by syntactic, pragmatic, and discourse considerations that go beyond concept formation as such, it is not surprising to find that their high, nearly 'syntactic' degree of productivity contrasts sharply with the semi-productivity of reversative *un-*, for which only concept formation is involved.

This is not to say that concept formation and semantics have no role to play for *-ly, -ness* and *un-*. For instance, of the affixes studied here, affix generalization occurs mainly with formations in *-ness* and *-ity*, even though these suffixes are not as productive as adverbial *-ly*. We have argued that the abstract semantics underlying the formation of these de-adjectival nominalizations apparently favor affix generalization. A high degree of productivity as such is not the driving force. The importance of the semantic function expressed by an affix is also apparent in a comparison of 'The Times' with the Cobuild/CELEX database. The greatest divergence between the two corpora emerges for *un-* and *in-*, suggesting that the productivity of affixal negation as such is a function of text type rather than the degrees of productivity of

the prefix forms *un-* and *in-* themselves.

The most productive affixes considered in this study give rise to lexical innovations that occur in increasing numbers over the months of sampling. This may be due to an increase in productivity of these affixes over time in the newspaper English of 'The Times', and perhaps in the English language. Alternatively, this increase in productivity may be due to changes in the population of journalists writing for 'The Times', or to changes in editorial standards. Whatever the explanation of the observed increasing rate of lexical innovation in 'The Times' may be, it would seem that readers of 'The Times' encounter some three de-adjectival formations each day that occur something like once every three years. As it is highly unlikely that these complex words are stored in the mental lexicons of individual speakers, our counts strongly suggest that, at least in reading, productive word formation rules are put to use on a regular daily basis.

Finally, we discussed empirical evidence in support of the hypothesis that the degree of productivity of an affix varies significantly with the morphological structure of the base word it attaches to. This issue, which has been hotly debated in the linguistic literature (Aronoff, 1976, Van Marle, 1985, Van Santen, 1992), can be laid to rest on the basis of the greatly varying

numbers of neologisms that we have observed in 'The Times'.

Author note.

We are indebted to Mark Aronoff, Beth Levin, Shelly Lieber, Robert Schreuder, and Richard Sproat for their valuable comments on earlier versions of this paper.

1. This research unit was at the University of Birmingham at the time of the AVIATOR project. AVIATOR was a research project funded jointly by the UK government and industry within the SALT Programme, with Harper Collins publishers, BRS Software and the University of Birmingham as its collaborating partners.

2. Use of the 'ordinary words' section of our database implies that capitalized sentence-initial words have not been taken into account. All words in the 'ordinary words' section of the database ending in one of the five affixes studied here were manually corrected for typing errors before further analysis.

3. The smoothed curves in the upper right and bottom panels of Figure 1 were obtained by means of non-parametric interpolation from the grouped

frequency distributions. This interpolation technique is based on the urn model for word frequency distributions. For the validity of the assumptions underlying this technique when applied to the frequency distributions of morphological categories, see Baayen (1994c).

4. Strictly speaking, *-ness* and *-ity*, and *un-* and *in-* are not fully synonymous, as subtle differences in meaning as well as differences in register and style characterize the use of these rival forms (see, e.g., Clark, 1993). We will nevertheless refer to these affixes as 'rival' because they are as similar as possible in meaning and use, and therefore allow us to investigate differences in productivity while controlling for other factors as much as possible.

5. In general, native speaker intuitions can be relied on with respect to judgements concerning the grammaticality of a complex word. For instance, most speakers of English would agree that *greenal* is not a possible word of English, in contrast to words such as *expressionlessness, tielessness* and *pennilessness*, all of which are clearly grammatical. However, intuitions concerning the familiarity of complex words should be handled with caution. Whereas speakers have clear intuitions about their familiarity with

low-frequency monomorphemic words such as *bawbee, clew, corm* and *flivver*, judgements of the familiarity of complex words such as *expressionlessness, tielessness* and *pennilessness* are heavily influenced by the familiarity of the base words *expression, expressionless, tie* and *tieless*, *penny* and *penniless*. Clearly, native speaker intuitions are not a reliable source of information for investigating whether low-frequency words are lexical innovations.

6. Note that the simple existence of large numbers of hapax legomena, representing one fifth up to one third of the total number of different word types (see Table 1), shows that the use of a lexical innovation should not be equated with the intentional introduction of a new term that is first defined and subsequently referred to. If that were the normal way for lexical innovations to enter the language, hapax legomena should be exceptional and less frequent than dis legomena and words with higher frequencies of use. Since the hapax legomena substantially outnumber the higher frequencies of use, term formation as such is unlikely to be the primary function of the ephemeral use of neologisms we observe in 'The Times'.

7. Although the overall degree of productivity of word formation in spo-

ken language is much lower than in written language (see section 2), the reverse may hold for the process of affix generalization observed for *-ness* and *-ity*.

8. Marchand (1969:205) argues that *un-* does not attach to non-resultative, durative verbs such as *smoke, wait* and *walk*. Exceptional formations such as *unlove* suggest that when this restriction is violated, the result is a verb with a terminative instead of a reversative reading.

9. In all counts presented here, including the comparison with the CELEX database to be presented in the next section, we have focused exclusively on the most productive allomorph of *in-*, *in-* itself, leaving formations with *ir-* and *im-* out of consideration. In the CELEX lexical database, the allomorph *in-* accounts for 77% of all types and 80% of all hapax legomena. Hence inclusion of the allomorphs is not expected to change the pattern of results obtained for 'The Times'.

10. Kjellmer (1984) proposes to measure the productivity of adverbial *-ly* with respect to a base type *X-x* in terms of the ratio of the number of for-

mations *X-x-ly* to the number of formations *X-x*. Note that this amounts to comparing the productivity of *-ly* with the productivity of the affixes occurring in its base words. Not surprisingly, counterintuitive results are obtained, such as that *-ly* would be more productive for base words in *X-ful* than for base words in *-al*, even though his dictionary-based counts of formations in *-ly* suggest a roughly similar pattern to the one we have observed for 'The Times', a pattern with relatively few words in *X-ful-ly* but large numbers of words in *X-al-ly*.

11. Williams' (1965: 285) critique of purist claims about proper English word formation mentions the possibility that the low degree of productivity of *-ness* with participial forms is due to the circumstance that it 'is likely to evoke condemnation, largely on 'aesthetic' grounds.'

12. Kjellmer (1984) argues that adverbial *-ly* is especially productive for dynamic adjectives, and that the degree of productivity of words in *-x-ly* is determined by the dynamism of affix *-x*. Unfortunately, 'static' base words

in *-al* appear to be more productive than 'dynamic' base words in *-ing*.

13. An earlier quantitative study of English derivational morphology that made use of the Cobuild/CELEX frequencies (Baayen and Lieber, 1991) suggested that *in-* is fairly productive, with a degree of productivity (measured with the $\mathcal{P}$ statistic) that is only slightly lower than that of *un-*. In fact, *in-* appeared in a larger number of different types than *un-*. There are two reasons why this unexpected result was obtained. First, as the above comparison of the Cobuild/CELEX database with our database of 'The Times' shows, *in-* is more productive in the former than in the latter. More importantly, but, unfortunately unknown to those authors, the CELEX database listed the frequency of occurrence only for a subset of words in the Cobuild Corpus, namely, the words occurring both in a dictionary-derived word list and in the Cobuild Corpus, and high-frequency words in the Cobuild Corpus not in that list. Consequently, large numbers of low-frequency words were not available. The counts of the Cobuild Corpus underlying figure 8 are based on the complete corpus, and clearly show that *un-* is much more productive than *in-*. Although in general the dictionary-conditioned corpus-based analyses in Baayen and Lieber (1991) generally appear to have captured the main trends,

the counter-intuitive results obtained for *un-* and *in-* are now resolved. From a methodological point of view, this is yet another illustration of the dangers inherent in a dictionary-based approach to morphological productivity.

References

Aronoff, Mark. 1976. Word formation in generative grammar. Cambridge, Mass.: The MIT Press.

Aronoff, Mark. 1980. The relevance of productivity in a synchronic description of word formation. Historical Morphology, ed. by Jacek Fisiak, 71–82. The Hague: Mouton.

Aronoff, Mark. 1982. Potential words, actual words, productivity and frequency. Preprints of the Plenary Session Papers of the XIIIth International Congress of Linguists, Tokyo, 141–48.

Baayen, R. Harald. 1992. Quantitative aspects of morphological productivity, Yearbook of Morphology 1991, ed. by Geert E. Booij and Jaap Van Marle, 109–49. Dordrecht: Kluwer Academic Publishers.

Baayen, R. Harald. 1993a. On frequency, transparency, and productivity. Yearbook of Morphology 1992, ed. by Geert E. Booij and Jaap Van Marle, 181–208. Dordrecht: Kluwer Academic Publishers.

Baayen, R. Harald. 1993b. Statistical models for word frequency distributions: A linguistic evaluation. Computers and the Humanities 26.347–63.

Baayen, R. Harald. 1994a. Derivational productivity and text typology. Journal of Quantitative Linguistics 1.16–34.

Baayen, R. Harald. 1994b. Productivity in production. Language and Cognitive Processes 9.447–69.

Baayen, R. Harald. 1994c. The randomness assumption in word frequency statistics. From Universal Morphisms to Megabytes: A Baayen Space Odyssey, ed. by Krzysztof Apt, Lex Schrijver and Nico Temme, 125-40. Amsterdam: Centre for Mathematics and Computer Science.

Baayen, R. Harald and Rochelle Lieber. 1991. Productivity and English derivation: a corpus-based study. Linguistics 29.801–43.

Baayen, R. Harald, Richard Piepenbrock, and Hedderik Van Rijn. 1993. The CELEX lexical database (CD-ROM). Philadelphia, PA.: Linguistic Data Consortium, University of Pennsylvania.

Bauer, Laurie. 1992. Scalar productivity and *-lily* adverbs. Yearbook of Morphology 1991, ed. by Geert E. Booij and Jaap Van Marle, 185–191. Dordrecht: Kluwer Academic Publishers.

Biber, Douglas. 1989. A typology of English texts. Linguistics 27.3–43.

Biber, Douglas and Edward Finegan. 1989. Drift and evolution of English style: A history of three genres. Language 65.487–517.

Bolinger, Dwight L. 1948. On defining the morpheme. Forms of English, accent, morpheme, order, ed. by Dwight L. Bolinger, 183–89. Cambridge, Mass.: Harvard University Press.

Bowerman, Melissa. 1982. Reorganizational processes in lexical and syntactic development. Language acquisition. The state of the art, ed. by Eric Wanner and Lila R. Gleitman, 319–46. Cambridge: Cambridge University Press.

Burrows, John. F. 1992. Computers and the study of literature. Computers and Written Texts, ed. by Christopher. S. Butler, 167–204. Oxford: Blackwell.

Butterworth, Brian. 1983. Lexical representation. Language production (Vol. II): Development, writing and other language processes, ed. by Brian Butterworth, 257–94. London: Academic Press.

Cannon, Garland. 1988. Historical change and English word-formation. New York: Peter Lang.

Chitashvili, Revas J. and R. Harald Baayen. 1993. Word frequency distributions. Quantitative Text Analysis, ed. by Gabriel Altmann and Luděk Hřebíček, 54–135. Trier: Wissenschaftlicher Verlag Trier.

Clark, Eve V. 1993. The lexicon in acquisition. Cambridge: Cambridge University Press.

Clark, Herbert H. and Eve V. Clark. 1979. When nouns surface as verbs. Language 55.567–811.

Cleveland, W. S. 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74.829–36.

Frauenfelder, Uli H. and Robert Schreuder. 1992. Constraining psycholinguistic models of morphological processing and representation: The role of productivity. Yearbook of Morphology 1991, ed. by Geert E. Booij and Jaap Van Marle, 165–83. Dordrecht: Kluwer Academic Publishers.

Good, I. J. 1953. The population frequencies of species and the estimation of population parameters. Biometrika 40.237–64.

Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. Meaning, form, and use in context:

Linguistic applications, ed. by Deborah Schiffrin, 11–42. Washington, D.C.: Georgetown University Press.

Kastovsky, Dieter. 1986. Productivity in word formation. Linguistics 24.585–600.

Kjellmer, G. 1984. Why *great:greatly* but not *big:bigly*? on the formation of english adverbs in *-ly*. Studia Linguistica 38.1–19.

Landauer, Thomas K. 1986. How much do people remember? Some estimates of the quantity of learned information in long-term memory. Cognitive Science 10.477–93.

Levelt, Willem J. M. 1989. Speaking. From intention to articulation. Cambridge, Mass.: The MIT Press.

Levin, Beth. 1993. English Verb Classes and Alternations. A preliminary Investigation. Chicago: The University of Chicago Press.

Marchand, Hans. 1969. The Categories and Types of Present-Day English Word Formation. A Synchronic-Diachronic Approach. München: Beck'sche Verlagsbuchhandlung.

Melville, Herman. 1851. Moby Dick. Project Gutenberg. Electronic version prepared by E.F.Tray, University of Colorado, Boulder, on the basis of the Hendricks House edition. Ftp-address: mrcnext.cso.uiuc.edu.

Merriam, George, and Charles Merriam. 1961. Webster's Third International Dictionary of the English Language. Addenda Section 1981. Springfield, Mass.: Merriam Webster Inc.

OED. 1987. The Oxford English Dictionary on Compact Disk. Ft. Washington, Pa.: Tri Star Publishing.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. A comprehensive grammar of the English language. London: Longman.

Rainer, Franz. 1988. Towards a theory of blocking: the case of Italian and German quality nouns. Yearbook of Morphology I, ed. by Geert E. Booij and Jaap Van Marle, 155–85. Dordrecht: Foris.

Renouf, Antoinette J. 1987. Corpus development. Looking up: An account of the Cobuild Project in Lexical Computing, ed. by John M. Sinclair, 1–40. London: Collins.

Renouf, Antoinette J. 1993a. Sticking to the text: A corpus linguist's view of language. Aslib Proceedings 45.131–36.

Renouf, Antoinette J. 1993b. A word in time: First findings from the investigation of dynamic text. English language corpora: Design, analysis, and exploitation, ed. by Jan Aarts, Pieter De Haan, and Nelleke Oostdijk, 279-88. Amsterdam: Rodopi.

Rubenstein, Herbert and Irwin Pollack. 1963. Word predictability and intelligibility. Journal of Verbal Learning and Verbal Behavior 2.147–58.

Scarborough, Don L., Charles Cortese, and Hollis S. Scarborough. 1977. Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance 3.1–17.

Schultink, Henk. 1962. De morfologische valentie van het ongelede adjectief in modern Nederlands [The morphological valency of the simplex adjective in modern Dutch]. Den Haag: Van Goor & Zonen.

Van Marle, Jaap. 1985. On the Paradigmatic Dimension of Morphological Creativity. Dordrecht: Foris.

Van Santen, Ariane. 1992. Produktiviteit in taal en taalgebruik [Productivity in Language and Language Use]. PhD Thesis, State University of Leiden. Leiden.

Whaley, C. P. 1978. Word–nonword classification time. Journal of Verbal Learning and Verbal Behavior 17.143–154.

Whorf, Benjamin L. 1953. Language, thought, and reality: Selected writings of Benjamin Lee Whorf (edited and with an introduction by John B. Carroll). Cambridge, Mass.: The MIT Press.

Williams, Theodore. 1965. On the "-ness" peril. American Speech XL.279–86.

	# new types	$V_N(1)$	V_N
-ly	40	1098	4900
-ness	29	739	2027
un-	25	659	1672
-ity	11	280	1020
in-	0	48	243

Table 1: Number of new types in the last month of sampling, number of hapax legomena in the sample ($V_N(1)$), and the total number of different word types (V_N) for *-ly*, *-ness*, *-ity*, *un-* and *in-*.

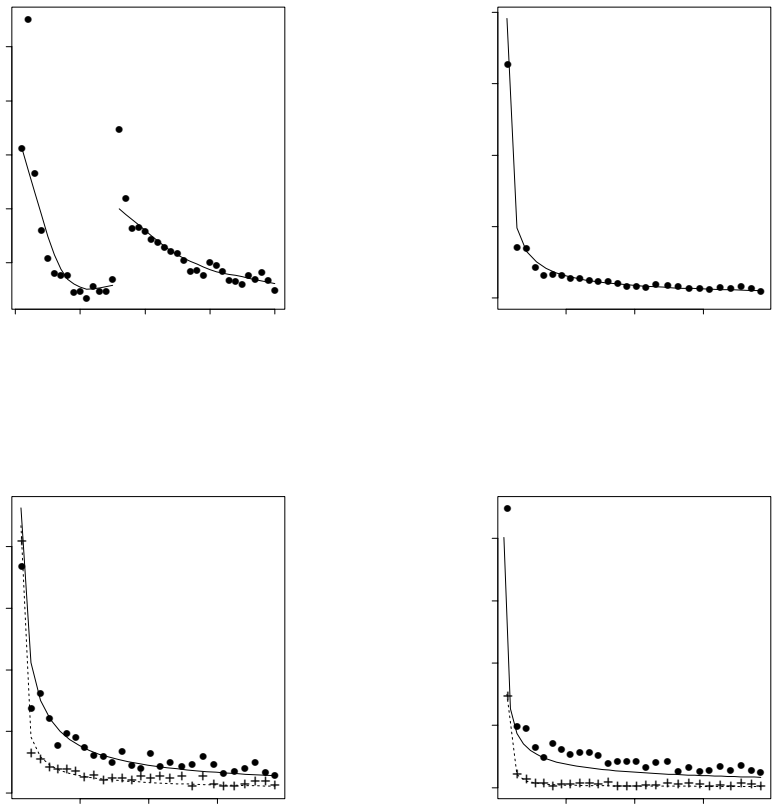

Figure 1: Numbers of new types as a function of sampling time (in months, top left panel, in number of tokens sampled, remaining panels) for *-ly*, top panels, for nominalizations in *-ness* (large dots) and *-ity* (+), bottom left, and for the negative prefixes *un-* (large dots) and *in-* (+), bottom right.

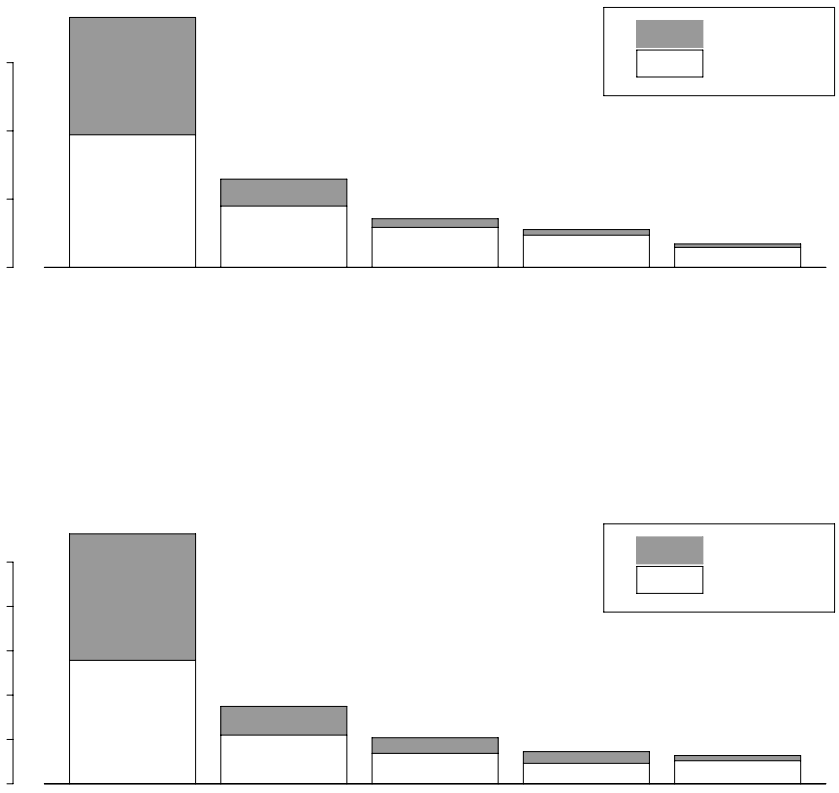

Figure 2: Numbers of types in *-ness* and -ity occurring in Merriam-Webster (1961/1981) dictionary (dict) and number of neologisms (neol) with respect to this dictionary for token frequencies 1–5.

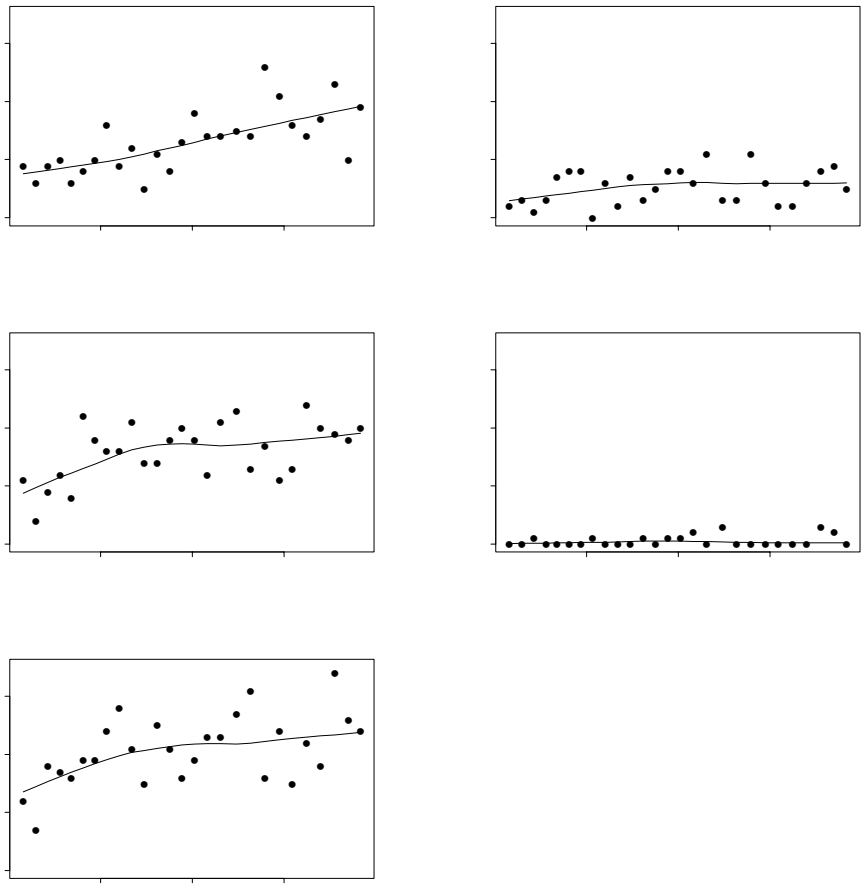

Figure 3: Numbers of new hapax legomena that are neologisms with respect to Merriam-Webster (1961/1981) as a function of the number of tokens sampled, for the affixes *in-* and *un-*, *-ness* and *-ity*, and *-ly*.

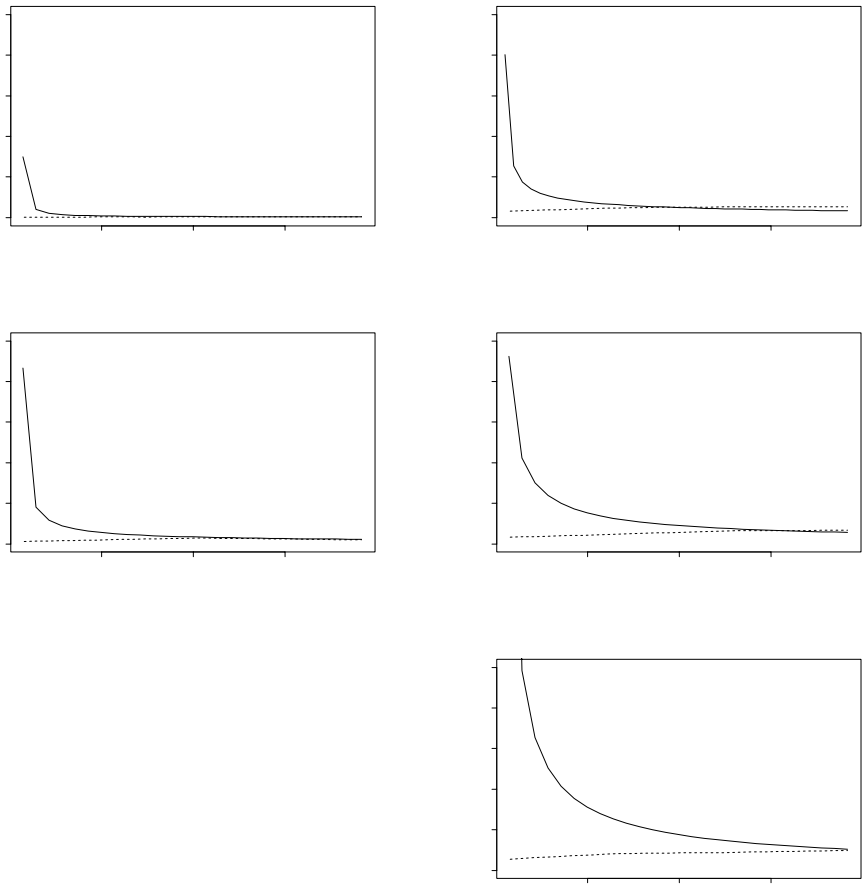

Figure 4: Numbers of new types (solid line) and number of new types occurring once only in the complete corpus (dotted line) as functions of the number of tokens sampled, for the affixes *in-* and *un-*, *-ness* and *-ity*, and *-ly*.

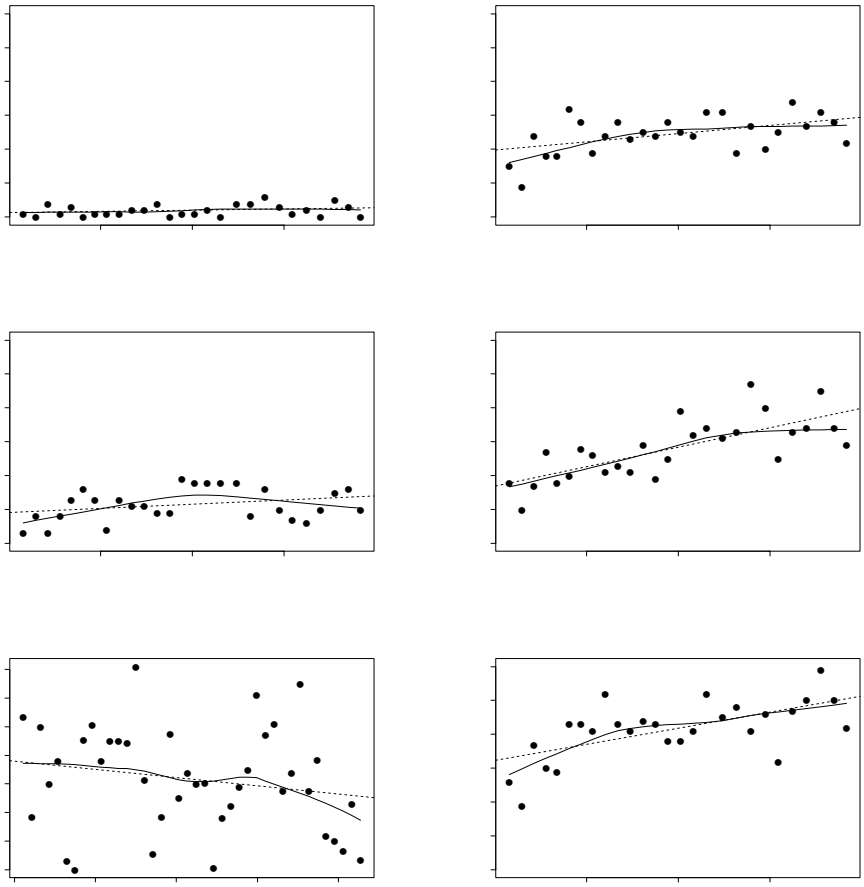

Figure 5: Number of new types occurring once only in the complete corpus as a function of the number of tokens sampled, for the affixes *in-* and *un-*, *-ness* and *-ity*, and *-ly*. The bottom left panel displays the distribution of hapax legomena in a novel, H. Melville's *Moby Dick*.

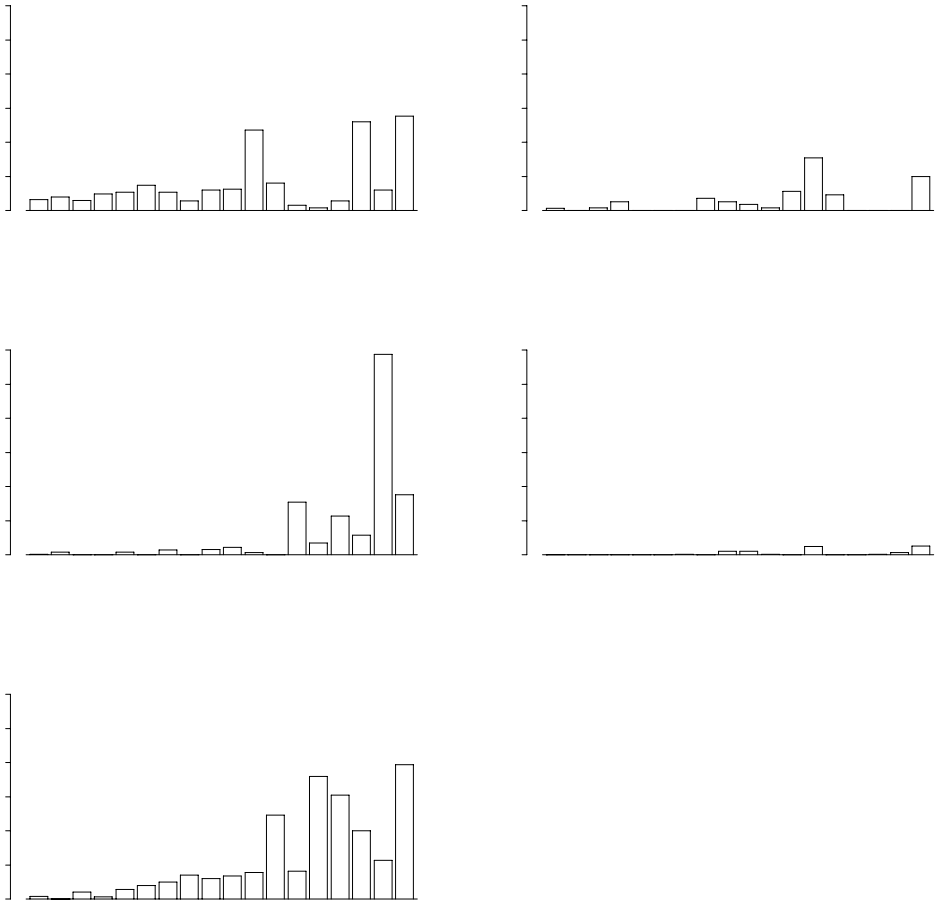

Figure 6: Numbers of hapax legomena in *-ness*, *-ity*, *un-*, *in*, and *-ity*, subdivided according to the kind of base word the affix is attached to (Phrase: phrasal base; AffG: affix generalization; Comp: compound; Mono: monomorphemic base word). The counts are based on the first 40 sampling months of the AVIATOR project.

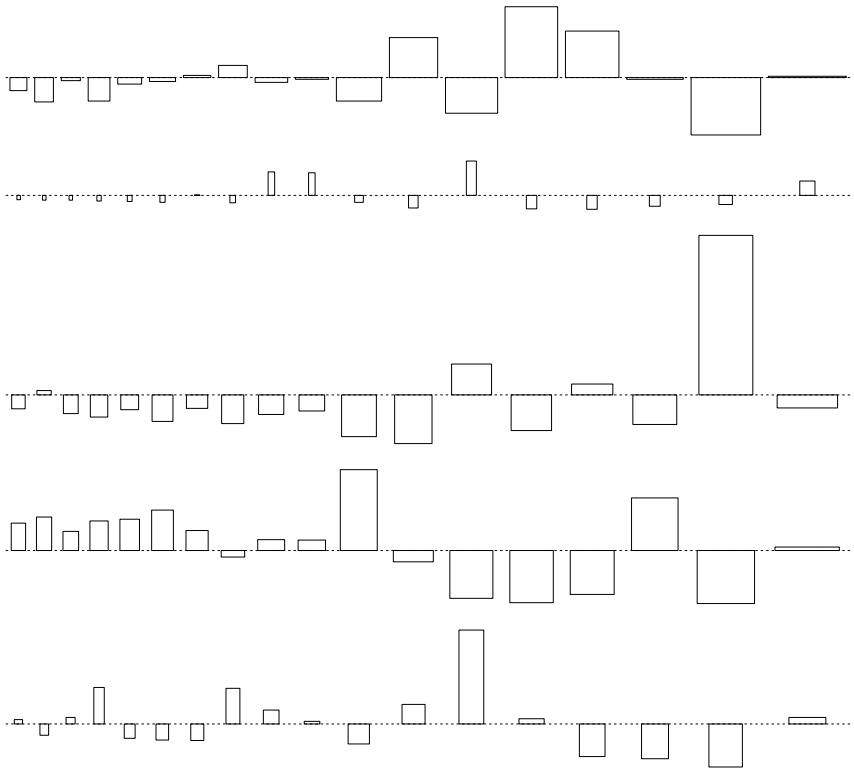

Figure 7: Chi-plot for loglinear fit to the number of hapax legomena as a function of affix and base word. (Squared) height proportional to the signed contribution to the Chi-squared Statistic, width proportional to the square root of the expected number of hapax legomena, area proportional to the residual deviance (the difference between the observed and expected number of hapax legomena). o: over-, A: affix generalization. The counts are based on the first 40 sampling months of the AVIATOR project.

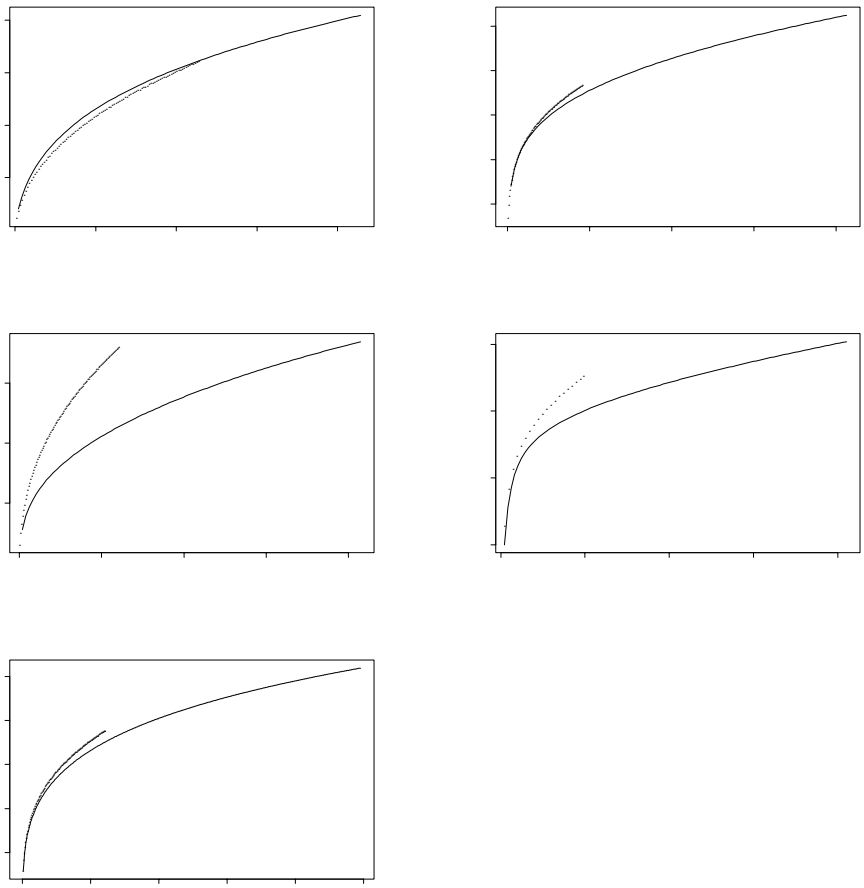

Figure 8: Interpolated vocabulary growth curve for the affixes *in-* and *un-*, *-ness* and *-ity*, and *-ly* for the Times corpus (solid line), and the Cobuild Corpus frequency data in the CELEX lexical database (dotted line).