

Suffix ordering and morphological processing

Ingo Plag and Harald Baayen

December 2, 2008

Abstract

There is a long-standing debate about the principles constraining the combinatorial properties of affixes. Hay (2002) and Hay and Plag (2004) proposed a model in which suffixes can be ordered along a hierarchy of processing complexity. We show that their model generalizes to a larger set of suffixes, and we provide independent evidence supporting their claim that a higher rank in the ordering correlates with increased productivity. Behavioral data from lexical decision and word naming show, however, that their model has been one-sided in its exclusive focus on the importance of constituent-driven processing, and that it requires supplementation by a second and equally important focus on the role of memory. Finally, using concepts from graph theory it is shown that the space of existing suffix combinations can be conceptualized as a directed graph, which, with surprisingly few exceptions, is acyclic. This acyclicity is hypothesized to be functional for lexical processing.¹

1 Introduction

In English and many other languages with derivational morphology, there are severe restrictions on possible combinations of affixes and bases. A given derivational affix attaches only to bases that have certain phonological, morphological, semantic, or syntactic properties. For example, the verbal suffix *-ize* only occurs on nouns and adjectives that end in an unstressed syllable, see Plag (1999) for details. Similar, or even more complex, restrictions seem to hold for affix-affix combinations. For instance, the word *atomic* can take the suffix *-ity* as a nominalizing suffix, whereas the word *atomless* can not take *-ity*, but can take the competing nominalizing suffix *-ness* (**atomlessity* vs. *atomlessness*).

There has been a long debate about whether there are general principles or mechanisms that constrain the combinatorial properties of affixes. In this debate three approaches can be distinguished. First, there are stratum-oriented models (e.g., Siegel 1974, Allen 1978, Selkirk 1982, Kiparsky 1982, Mohanan 1986, Giegerich 1999) that claim that the lexicon has

¹The order of the authors is in reverse alphabetical order. The authors are indebted to Jen Hay, the anonymous reviewers and the editors for their constructive criticism of earlier versions of this paper, as well as to the audiences at colloquia and workshops in Nijmegen, Mainz, Paris, Provo and Vienna for stimulating discussion.

a layered structure and that this structure largely determines the combinatorial properties of affixes. Second, there are scholars who argue that affix-particular selectional restrictions (of a phonological, morphological, semantic or syntactic nature) are responsible for possible and impossible combinations of affixes (e.g. Fabb 1988, Plag 1999).

Most recently, a third theory has been proposed in Hay (2003), which holds that constraints on the processing of morphological structure control affix combinations. In Hay & Plag (2004), the three models were empirically tested against the 210 potential two-suffix combinations of 15 English derivational suffixes (most of them Germanic 'stratum 2' suffixes). They observed that both selectional restrictions and processing constraints are instrumental in determining possible and impossible suffix combinations. Only well processable combinations are possible combinations, and this range of possible combinations is further curtailed by selectional restrictions. In essence, the model proposed by Hay & Plag (2004) proved not only empirically superior to stratum-oriented models, but also provides a psycholinguistically attractive solution to the long-standing problems of suffix-suffix combinations.

The present paper extends Hay and Plag's research program in two important ways. After a review of relevant literature, we investigate a specific suffix combination in Hay and Plag's set of combinations for which their model apparently makes the wrong predictions (Section 3). We then submit their model to even more serious empirical challenges by doubling the set of English suffixes under investigation, which increases the number of combinations tested from 210 (in Hay & Plag 2004) to 930 (Section 4). Second, if Hay and Plag's theory of *Complexity-Based Ordering* is indeed psycholinguistically plausible, affix ordering should be predictive for processing complexity. Thus, we investigate whether the ranking of suffixes obtained with our enlarged data set not only correlates with distributional measures such as productivity, but also with experimental estimates of actual lexical processing costs as gauged by latencies in word naming and visual lexical decision tasks (Section 5). In section 6 we summarize our findings and discuss their theoretical implications.

2 Complexity-Based Ordering: Selectional and processing restrictions

Until a few years ago the debate on stacking restrictions was characterized by two opposing views. Proponents of stratum-oriented models (such as Siegel 1974, Allen 1978, Selkirk 1982, Kiparsky 1982, Mohanan 1986) assume that most, if not all combinatorial restrictions among English suffixes can be explained by the fact that these suffixes belong to different lexical strata and that these strata interact phonologically and morphologically in intricate ways. This is known as level-ordering, which in turn is part of most models of Lexical Phonology. According to the level-ordering hypothesis, English suffixes and prefixes belong to the following classes or strata (from Spencer 1991:79):

Class I suffixes: +ion, +ity, +y, +al, +ic, +ate, +ous, +ive, +able, +ize

Class I prefixes: re+, con+, de+, sub+, pre+, in+, en+, be+

Class II suffixes: #ness, #less, #hood, #ful, #ly, #y, #like, #ist, #able, #ize
Class II prefixes: re#, sub#, un#, non#, de#, semi#, anti#

The suffixes belonging to one stratum are said to share a number of properties that distinguish them from the suffixes of the other stratum. Stratum 1 suffixes tend to be of foreign origin ('Latinate'), while stratum 2 suffixes are mostly Germanic. Stratum 1 suffixes frequently attach to bound roots and tend to be phonologically and semantically less transparent than stratum 2 suffixes. Stratum 1 suffixes cause stress shifts, resyllabification, and other morpho-phonological alternations, stratum 2 suffixes do not. Stratum 1 suffixes are less productive than stratum 2 suffixes, and, crucially, stratum 1 suffixes do not occur outside stratum 2 suffixes. Thus, suffixes can only combine in such a way that they attach to suffixes of the same stratum or of a lower stratum. This is perhaps the most important generalization concerning suffix combinations that emerges from stratum models, since combinations in which a class II suffix occurs inside a class I suffix (such as **atom#less+ity*) are ruled out on principled grounds.

However, there are serious problems with this approach. One major theoretical weakness of level ordering is that the two strata are not justified on independent grounds. In other words, it is unclear what is behind the distinction between the two strata, and which properties make a suffix end up on a given stratum. The idea that the underlying distinction is one of etymology (borrowed vs. native, e.g. Saciuk (1969)) does not explain why speakers can and do master English morphology with little or no etymological knowledge. Others have argued that the stratum distinction is phonological in nature, with differences between different etymological strata being paralleled by phonological differences (see, e.g., Booij (2002), Van Heuven et al. (1994) for Dutch). Although this would allow speakers to distinguish between the strata on the basis of the segmental and prosodic behavior of derivatives, a phonological explanation of the nature of the strata would weaken the idea of strata considerably. As Raffelsiefen (1999) shows, not even two of the many suffixes of English trigger exactly the same type of morpho-phonological alternations, so that we would need as many sub-strata as we have suffixes that trigger morphological alternations. Thus we end up with a continuum, rather than with a discrete two-level system.

Another serious problem is that a stratum cannot either be defined by the set of suffixes it contains, because many suffixes belong to more than one stratum, given that in certain derivatives they show stratum 1 behavior, whereas in other derivatives they display stratum 2 behavior, with sometimes even doublets occurring (cf. *compárable* vs. *cómparable*). Furthermore, there are a number of unexpected suffix combinations. Thus stress-neutral *-ist* appears inside stress-shifting *-ic*, or stress-neutral *-ize* appears inside stress-shifting *-(at)ion*. In order for the model not to make wrong predictions, dual membership of affixes (or some other device weakening the overall model) becomes a necessity.

Giegerich (1999) discusses cases of apparent dual membership of affixes in great detail and, as a consequence, proposes a thoroughly revised stratal model, in which the strata are no longer defined by the affixes of that stratum, but by the bases. This base-driven stratification model, which is enriched by many suffix-particular base-driven restrictions, can overcome some inadequacies of earlier stratal models, but at the cost of significantly reduc-

ing the overall predictive power of the model. These restrictions are a well-taken step towards eliminating the weakness of not making any predictions concerning suffix order within strata, which characterized earlier Lexical Phonology models. Important problems remain, however. For example, Fabb (1988) and Plag (1996, 1999) point out that there are numerous other important (phonological, morphological, semantic, syntactic) restrictions operative in English suffixation. Level ordering does not say anything about these restrictions. For example, Fabb finds that the 43 suffixes he investigates are attested in only 50 combinations, although stratum restrictions would allow 459 out of the 1849 possible ones.

In general, for any given affix, its phonological, morphological, semantic and syntactic properties (or the properties of its derivatives, i.e. of the morphological category) must be stated in its lexical entry. Plag (1996, 1999) shows that these diverse properties together are responsible for the combinatorial potential of a given affix. What has been analyzed as would-be stratal behavior thus falls out from the phonological, morphological and semantic properties of the affix. Since these properties must be stated anyway to account for the particular behavior of a given affix, no further stratal apparatus is necessary.

What is the alternative? Hay & Plag (2004) propose that selectional restrictions and processing constraints together are responsible for the attested and non-attested patterns of suffix-suffix combinations. Selectional restrictions are understood as affix-particular properties which govern the kinds of combinations that are allowed for that affix. Such restrictions can refer to phonological, morphological, syntactic or semantic characteristics of the elements to be combined. For example, verb-forming *-en* (as in *blacken*) only attaches to monosyllables that end in an obstruent.² This phonological restriction also has immediate consequences for suffix combinations with *-en* as the outer suffix. Given that *-en* attaches only to monosyllables, *-en* may never attach to bases that themselves contain a suffix that creates a new syllable, which would be the case for most adjectival suffixes (such as *-al*, *-ive*, *-ous*, etc.). We can therefore predict that the combinations *-al-en*, *-ive-en*, and *-ous-en* will not occur, due to the selectional restriction of *-en*. Another example, this time involving a morphological restriction, is the suffix *-ize*, which selects only *-ation* as the nominalizing suffix (as in *colonization*). An instance of a semantic restriction is the verbalizing suffix *-ate*, which attaches productively only to nouns that refer to chemical substances (such as *fluorinate* Plag 1999). Since such substances are usually denoted by nouns, *-ate* will never follow an adjectival or verbal suffix, or an abstract noun-forming suffix such as *-ism*.

In addition to selectional restrictions, Hay (2002, 2003) proposed a general processing constraint, named *Complexity-Based Ordering* by Plag (2002), that limits the number of possible combinations of affixes. The underlying idea is that the morphological separability of affix and base is a graded phenomenon, with far-reaching consequences for affix stacking:

While some affixes basically tolerate no internal structure, others will tolerate structure to some minimum degree. The degree of internal structure tolerated by an affix is ... determined by how much structure that affix, itself, creates.

²If defined over roots, this restriction even holds for the parasynthetic formations in which *-en* occurs, for example, *enlighten*, *embolden*.

Phrased in terms of processing, *an affix which can be easily parsed out should not occur inside an affix which can not.* (Hay, 2002:527–528, our emphasis)

To understand this better, we must take a closer look at morphological processing. It is often argued that there are two ways of processing a morphologically complex word. Either the word is decomposed, in which case it is produced or understood through its constituents (*in-* and *sane* for *insane*) or processed as a whole (*insane*), as if it were a monomorphemic form. In parallel dual route models of morphological processing, whole-word access (critically depending on storage) and access through constituents (critically depending on rules) are used simultaneously. Which of the access routes is most efficient for a given word depends on the frequency of the derived word, the frequencies of its constituents, and various other processing parameters relating to word length, lexical competition, and semantic connectivity (Baayen & Schreuder 1999, Baayen & Moscoso del Prado Martín 2005).

Hay (2001) has argued that the preference for one of the routes depends specifically on *relative frequency*, i.e., the relation between the frequency of the derived word and that of its base. Hay (2001) shows that the higher the frequency of the derived word in relation to the base word, the less likely is constituent-based processing. Alternatively, the lower the frequency of the derived word in relation to the base word, the more important the role of the constituents becomes. Let us look at an example. The derived word *government* (BNC lemma frequency 66894) is much more frequent than its base *govern* (BNC lemma frequency 2626), hence there is a whole-word bias for *government*. Note that this whole-word bias is reflected in the pronunciation of *government*, which involves either assimilation of base-final /n/ to the suffix-initial /m/, or even the complete loss of the last syllable of the base. Such phonological opacity is typical of lexicalized forms, i.e. forms with a whole-word bias in processing. In contrast, *discernment* (BNC 61) is much less frequent than its base *discern* (BNC 452), which leads to a strong bias for constituent-driven processing.

Such a view on complex words has serious implications for the notion of morphological complexity. First, it means that the role of the same suffix in lexical access will be different in different words depending on the respective frequencies of base and derivative (e.g., *-ment* plays a greater role for *discernment* than for *government*). Second, suffixes occurring in many words which are less frequent than their bases will tend to be more important in processing than suffixes represented by few words which are less frequent than their bases (compare, for instance, *-ness* and *-ic*). Finally, affixes that are stronger in processing will occur outside affixes that are weaker, an idea already expressed in Burzio, 1994: 354).

Hay & Plag (2004) show that suffixes can indeed be ordered in a hierarchy of juncture strength, such that affixes following an affix A on the hierarchy can be added to words already containing A, but affixes preceding A on the hierarchy cannot freely attach to words containing A. Thus, given a hierarchy of suffixes X-Y-Z-A-B-C-D, possible combinations would, for example, be *base-A-B*, *base-X-A-C*, or *base-Y-Z-A*, whereas **base-A-Z*, **base-Y-A-Z*, and **base-X-A-Y* would be impossible combinations.

Hay and Plag investigated the relationship between processing constraints and selectional restrictions (grammatical constraints) by looking at attested and non-attested combinations of 15 English suffixes (*-dom*, *-ee*, *-en*, *-er*, *-ess*, *-ful* (adjectival), *-ful* (nominal), *-hood*, *-ish*,

-less, -ling, -ly, -ness, -ship, -th). They checked all possible combinations ($N = 210$) for attestations in the British National Corpus (BNC, Burnard 1995), the CELEX lexical database (Baayen et al. 1995), and the Oxford English Dictionary (OED 1994), supplemented by some data obtained through internet searches, and found that the attested combinations can be arranged in a strict hierarchy. This hierarchy of suffixes established on the basis of attested combinations correlates with the order of suffixes established on the basis of suffixal processing measures, such as relative frequency and productivity, as computed in Hay & Baayen (2002). This correlation provides evidence for the idea that the hierarchy is largely constrained by processing. In addition to such processing considerations, Hay and Plag show, however, that the non-attested combinations are also largely excluded by selectional restrictions and that the range of combinations allowed by the general restrictions on constituent-driven processing is further curtailed by selectional restrictions. The authors conclude that processing constraints and grammatical constraints (selectional restrictions) work hand in hand, such that the grammar tends to create structures that are better processable. Similar insights can be found in Hawkins (2004) for word order, and in Bybee (1985) for affix order from a typological perspective.

The model proposed by Hay & Plag (2004) raises some problems and questions. A first problem, first pointed out by Plag (2002), is that there are suffixes that quite regularly appear inside and outside of each other. For example, adjectival *-al* may appear inside and outside of nominal *-ion*, as illustrated in *sens-ation-al* and *coloni-al-iz-ation*. In Hay & Plag (2004) these cases are also considered and it is shown that in such cases of multiple affixation, a middle suffix such as *-ize* preferably attaches to *-al* derivatives with a low level of decomposability. The same would hold for individual formations with a whole word processing bias, such as *business*, which could in principle be the base for suffixes that are lower in the hierarchy than *-ness*. What may be surprising is the fact that very few of such reverse combinations are actually attested. This is understandable, however, in the light of two facts. First, such a reverse order would still have to observe the selectional restrictions of the suffixes involved. This severely restricts the range of possible combinations. Second, outer suffixes are outer suffixes exactly because most of their derivatives are better parsable than inner suffixes, which means that we find relatively few derivatives with outer suffixes that have a whole-word bias. This in turn reduces the probability of new formations based on whole word bias forms with outer suffixes.

A related problem concerns the potential combinations of *-ness* and *-less*. Both combinations seem to be structurally possible in the sense that they are allowed by the selectional restrictions of these suffixes. Interestingly, Hay and Plag found a number of forms in their data involving the combination *-less-ness* (as in *hopelessness*), but did not find the combination *-ness-less* (as in the invented *happinessless*). This is all the more surprising because *-less* is more productive than *-ness* according to the measures developed in Hay and Baayen, and therefore should preferably occur outside of less productive *-ness*. The theory would therefore predict, contrary to fact, that formations such as *happi-ness-less* should be in actual use, rather than formations such as *hope-less-ness*. It is therefore necessary to investigate whether this mismatch between theory and data is due to the limitations of the database

used by Hay and Plag.

A second problem of Hay & Plag (2004) which needs to be addressed is that the set of suffixes comprised only 15 forms, most of which are stratum 2 suffixes. This raises the question of whether their account would also hold for a wider range of suffixes, including both stratum 1 and stratum 2 suffixes.

A third problem concerns the indirect nature of the evidence supporting the relationship between processing costs on the one hand and productivity and relative frequency on the other. The theory would gain weight if the current indirect evidence could be supplemented by more straightforward experimental evidence. Without experimental support, claims about psychological plausibility remain hypothetical. In the following sections, these three problems are addressed in turn.

3 The inside-outside problem: *-ness-less* vs. *-less-ness*

Is the absence of formations ending in *-ness-less* in the data surveyed by Hay and Plag due to limitations of the size of their data base (CELEX, BNC and OED)? Arguably the world-wide web provides the largest current resource of English usage, and we should therefore be able to trace such formations in this resource, if they exist. To this end, we took all 2466 attested *-ness* words from the BNC (Burnard 1995), added to them the string *less* and carried out a systematic internet search. We used the software made available by Hayes (2001) to search for pertinent words via Google on English language webpages. The search yielded quite a number of attestations and the results thus confirm the prediction of the structural analysis that *-ness-less* should be possible, given the selectional restrictions of these two suffixes. Table 1 lists the ten most frequently attested forms and their frequencies (as of November 17, 2007) for illustration.

Table 1: Combinations of *-ness-less* attested on the internet.

derivative	Google frequency
<i>happinessless</i>	303
<i>consciousnessless</i>	192
<i>businessless</i>	190
<i>thicknessless</i>	160
<i>randomnessless</i>	120
<i>sadnessless</i>	115
<i>lightnessless</i>	109
<i>weaknessless</i>	81
<i>darknessless</i>	78
<i>emptinessless</i>	69

A closer look at the individual websites reveals that many websites are counted more than once in Google and that many attestations are quite dubious, because they involve typos, word play, or are coined by non-native speakers. However, it is possible for each *-ness-less* formation listed in Table 1 to find forms that are highly natural and embedded in idiomatic native English. To illustrate this point, we give citations for the first four words in the list.

Eight Deadly Sins Of Web 2.0 Start-Ups . [...] **Happinessless**: Your start-up has no future if you are not happy <http://www.slideshare.net/imootee/eight-deadly-sins-of-web-20-startups/>

He walked past the sleeping fat man quietly, careful not to disturb something he himself envied. How long had it been since he'd slept, since he'd drifted into such a **consciousnessless** slumber? He got up and walked the corridors. <http://www.geocities.com/jacory69/SHORTFICTION.html>

"A lot of people are very alarmed, and this is only natural . . . but whether they are a home or a business, they are not going to be left homeless or **businessless**," Ed Weirauch, spokesman for the Camden Redevelopment Agency, said in a telephone interview yesterday. <http://www.saveardmorecoalition.org/node/887>

I'm trying to figure out if the general, conceptual skeletons of the Latin letter-forms (independent of a given design) need to have some information about the stroke thickness (certainly relative, not absolute), to be useful/make sense. Normally, if you ask somebody to draw the "essential" form of a letter (or visualize it in your head), you get a **thicknessless** diagram. <https://listserv.heanet.ie/cgi-bin/wa?A2=ind9904&L=typo-1&P=12980>

Interestingly, in three of the four examples the *-ness* derivative on which the *-ness-less* form is based is mentioned in the preceding discourse, and the previous mentioning seems to make the *-ness* form more easily available for further affixation (see, e.g., Kastovsky 1986, Baayen & Neijt 1997). This raises the question (Hay, personal communication) whether the *-ness* derivatives that take *-less* are a random selection from all *-ness* derivatives, or whether they are forms that are especially prone to further affixation, for example due to their frequential properties.

In order to check whether the bases in *-ness* that take *-less* are somewhat special (and not randomly chosen from all *-ness* derivatives), we checked for all ten *-ness* bases of the words in Table 1 their frequencies and relative frequencies in CELEX. According to Hay (2001), we should expect that these bases range among the most frequent *-ness* words, and that their relative frequencies are also rather high in comparison to all other *-ness* words. Both of these facts would make these words more prone to further affixation than *-ness* words that are less frequent or have a smaller relative frequency. In other words, the *-ness* derivatives in question should not be randomly selected.

Both of the above predictions are borne out by the facts. CELEX contains 1316 different *-ness* derivatives, with frequencies ranging from 4409 to 0 (0 pertains to data that were not found in the COBUILD corpus but entered CELEX through dictionaries). Of the six

most frequent *-ness* derivatives in CELEX, five feature as bases in Table 1 (*business* ranks first,³ *darkness* comes in second, *consciousness* fourth, *weakness* fifth, and *happiness* in sixth position). The least frequent base of the formations in Table 1, CELEX, *randomness*, still has rank 218 in CELEX. As expected, the frequencies of these forms correlate with their respective relative frequencies ($\rho = 0.95, p < 0.001$, Spearman’s rank correlation). The relative frequencies of the ten *-ness* bases occupy ranks between 24 and 451 among the 1316 *-ness* derivatives, which means that all forms in question can be found, roughly speaking, in the upper third of all *-ness* derivatives. In sum, our ten bases are not a random pick from all *-ness* derivatives. Their frequential properties make them especially prone to further affixation, in line with the assumptions of Complexity-Based Ordering (see also Krott et al. 1999).

Returning to the question of parsability of *-ness* vs. *-less*, we find that the suffix *-less* has higher figures than *-ness* on four measures of separability (Hay & Plag 2004): productivity: 0.016 versus 0.008; token parsing ratio: 0.74 versus 0.23; type parsing ratio: 0.86 versus 0.51; boundary strength (average rank according to the preceding three measures): 13.33 versus 8.67. Hence, *-less* should occur to the right of *-ness*, and we now see that forms instantiating the predicted order do indeed exist.

However, we are still left with the problem why these are obviously less common than the opposite, less easily parsable sequences. In fact, the problem of the two structurally possible combinations (*-less-ness* and *-ness-less*) is yet another instance of affix sets that seem to violate the predictions of Complexity-Based Ordering, similar to *-al*, *-ion* and *-ize*, mentioned above.

With our *-ness-less* and *-less-ness* derivatives, however, we have a somewhat different situation in that there are only two suffixes involved that can swap positions. This is predictable under the selectional restriction approach, but unexpected by Complexity-Based Ordering. Similar examples are discussed below. In the general discussion, we return to discuss this problem for Complexity-Based Ordering. In what follows, we first seek to broaden the empirical scope for the hypothesis of Complexity-Based Ordering.

4 Extending the data set to 31 suffixes

We extended Hay and Plag’s set of 15 suffixes by adding another 16 suffixes to test whether the predictions of Complexity Based Ordering are also borne out if more suffixes are taken into consideration, and if more suffixes from stratum 1 are also included. Table 2 lists all suffixes with their selectional restrictions. The last column shows whether the suffix has been part of Hay and Plag’s data set, or is newly introduced in this paper. The selectional

³The status of *business* as a derivative of the suffix *-ness* may be questionable, since the form is semantically and phonologically opaque. However, the suffix can still be phonologically discerned and the word has some properties that are typical of *-ness* formations, i.e. it is an abstract noun. As discussed in the previous section, the strength of the morphological boundary between a given affix and the different stems it attaches to can vary a great deal and is, among other things, dependent on the frequencies of base and derivative. Needless to say, *business* has a strong whole word bias, which, as in the case of *government*, goes together with its morphological opacity.

restrictions are derived from pertinent reference works on this question, such as Jespersen (1942), Marchand (1969), Bauer (1983), Adams (2001), Plag (2003), but also more specialized treatments on individual suffixes, such as Barker (1998), Ryder (1999), Dalton-Puffer & Plag (2000), Malkiel (1977), Riddle (1985), Ljung (1970), Plag (1999).

Overall, our data set amounts to 930 potential combinations, of which we found 160 attested in CELEX, the BNC or the OED. To this we added the combination of *-ness* followed by *-less* that we documented above. The attested 161 combinations and pertinent examples are documented in the Appendix. Table 3 shows the attested combinations, with the suffixes listed in alphabetical order. Note that the table shows a 1 even if there was only a single derivative of its kind attested in our vast data base. We deliberately did not impose a threshold on the number of attested forms, in order to increase the chance of falsifying the hierarchy hypothesis. The table can be read as follows: a 1 indicates that a combination is attested, with the suffix on the left margin as the inner suffix and the suffix on the upper margin as the outer suffix. In graph theory (see, e.g., Jungnickel 2007) Table 3 is referred to as the adjacency matrix of a directed graph.

We can see a great many zero cells, the vast majority of which are naturally explained by the work of selectional restrictions. Let us take a look at the first combinations in the next-to-last row for illustration. The suffix in the left margin, *-th*, can combine with *-en*, because it has monosyllabic derivatives (like *length*) and also meets the other requirements imposed by *-en* on its bases. The suffix *-th* may, however, not precede *-or*, *-ment*, or *-ive*, since these suffixes normally attach to verbs, and *-th* creates nouns. The combination *-th-aryN* is ruled out by the semantic restriction of *-aryN*, which takes only nouns as bases that denote persons, objects, or locations. Note also that the selectional restrictions are sometimes not clear enough to exclude certain combinations. For example, the combination *-th-ster* is not ruled out by any of the known selectional restrictions on bases of *-ster*, as listed in Table 2. In the vast majority of cases, however, an empty cell indicates that this combination is not only not attested, but that it is also structurally impossible. This finding is in line with Hay and Plag (2004), who also show for their data set that the suffix-particular selectional restrictions rule out the vast majority of combinations.

Returning to Complexity-based Ordering, its prediction now is that the suffixes can be re-ordered in this table in such a way that there are no affix combinations below the main diagonal of the adjacency matrix. In graph theory, this prediction is equivalent to the hypothesis that the adjacency matrix defines a directed acyclic graph. To understand this prediction and its implications, consider *-ess*. Table 3 shows us that this suffix attaches to words ending in, for instance, *-er*, *-ist* and *-ian*. We should therefore rank these three suffixes before *-ess* in the ordering hierarchy. The prediction now is that none of the suffixes that can attach to words ending in *-ess*, such as *-less*, *-dom* or *-ship* can precede *-er*, *-ist* and *-ian* in other derived words. Since *-less*, *-dom* and *-ship* can attach to words ending in *-ess*, they have to be ordered after *-ess*. If a suffix preceding *-ess* in the hierarchy would nevertheless precede *-dom*, *-ship* or *-less* in one or more complex words, than these affix combinations would show up below the main diagonal of the adjacency matrix. The left panel in Figure 1 illustrates the 15 combinations that are ruled out by Complexity-based

Table 2: Selectional restrictions of the 31 suffixes investigated in the present study. Question marks indicate unclear restrictions.

suffix	examples	derived category	base category	semantic restriction on base	semantic category of derivative	phonological restriction on base	data set
<i>-age</i>	<i>orphanage</i> <i>brewage</i>	N	N, V	concrete noun event	nomen actionis location	monosyllables(?)	new
<i>-ary</i> (N)	<i>apothecary salary</i>	N	N	person/object noun location			new
<i>-ary</i> (A)	<i>elementary</i> <i>arbitrary</i>	A	N		relational noun		new
<i>-(e)ry</i>	<i>slavery bakery</i> <i>monastery</i>	N	N, V		location collective noun status		new
<i>-ette</i>	<i>kitchenette</i> <i>leatherette</i> <i>suffragette</i>	N	N	concrete noun	diminutive		new
<i>-fold</i>	<i>tenfold manifold</i>	A	Numeral, many		qualitative adjective		new
<i>-(i)an</i>	<i>European</i>	N/A	N		relational/qualitative adjective		new
<i>-ism</i>	<i>patriotism</i> <i>positivism</i> <i>baptism</i>	N	N, A		abstract noun		new
<i>-ist</i>	<i>artist socialist</i> <i>baptist</i>	N	N, A		person noun		new
<i>-ive</i>	<i>active massive</i> <i>coercive</i>	A	V, N		relational/qualitative adjective		new
<i>-ment</i>	<i>amazement</i> <i>dreamment</i> <i>ornament</i>	N	A, V		nomen actionis		new
<i>-or</i>	<i>actor error</i>	N	V		person/instrument/ etc.		new
<i>-ous</i>	<i>courageous</i> <i>obvious</i>	A	N		relational/qualitative adjective		new
<i>-ster</i>	<i>gangster</i> <i>youngster</i>	N	N, A		person noun	monosyllables	new
<i>-wise</i> <i>-ly</i>	<i>lengthwise</i> <i>gladly</i>	Adv Adv	N A		adverb		new new

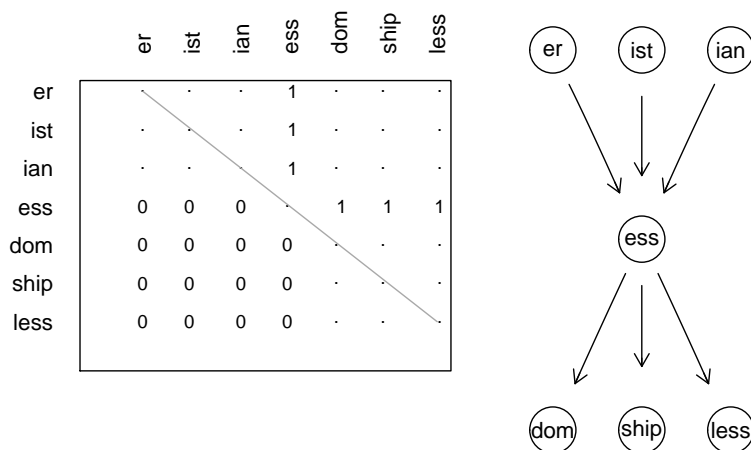
Table 2 (continued). Selectional restrictions of the 31 suffixes investigated in the present study.

suffix	examples	derived category	base category	semantic restriction on base	semantic category of derivative	phonological restriction on base	data set
-dom	<i>freedom</i>	N	N/(ADJ)	?	status, realm, collectivity ?		Hay & Plag 2004
	<i>stardom</i>	N	V/N	?	sentient being (non-agent)		Hay & Plag 2004
-ee	<i>employee</i>	V	ADJ/N	?	change-of-state verb	monosyllable, obstruent-final	Hay & Plag 2004
-en	<i>deepen</i>	N	V/N	?	person, instrument, etc.		Hay & Plag 2004
-er	<i>baker</i>	N	N	(male?)	female		Hay & Plag 2004
-ess	<i>princess</i>	ADJ	N	person/animal abstract noun	person/animal qualitative adjective		Hay & Plag 2004
-ful (ADJ)	<i>careful</i>	N	N	concrete noun	partitive noun		Hay & Plag 2004
	<i>cupful</i>	N	N/(ADJ)	person noun	state of being		Hay & Plag 2004
-ful (N)	<i>childhood</i>	ADJ/Num	N/ADJ/Num	N=person noun	X		Hay & Plag 2004
-hood	<i>falsehood</i>	N	N	animate noun	simulative	?	Hay & Plag 2004
-ish	<i>Jewish</i>	ADJ	N	?	without X		Hay & Plag 2004
	<i>schoolboyish</i>	N	N	animate noun	young animal, (young) human being	?	Hay & Plag 2004
	<i>greenish</i>	ADJ	N/(ADJ)	N=person and time nouns	simulative	?	Hay & Plag 2004
	<i>fortyish</i>	N	ADJ/(any)	?	quality noun		Hay & Plag 2004
-less	<i>careless</i>	N	N	person noun	status, collectivity		Hay & Plag 2004
-ling	<i>duckling</i>	ADJ	ADJ/V	?	quality noun	monosyllable	Hay & Plag 2004
-ly	<i>fatherly</i>	N	N	?	quality noun		Hay & Plag 2004
	<i>deadly</i>	N	ADJ/V	?	quality noun		Hay & Plag 2004
-ness	<i>kindness</i>	N	ADJ/(any)	?	status, collectivity		Hay & Plag 2004
-ship	<i>friendship</i>	N	N	person noun	quality noun		Hay & Plag 2004
-th	<i>depth</i>	N	ADJ/V	?	quality noun	monosyllable	Hay & Plag 2004

Table 3: Adjacency matrix for thirty-one suffixes and their pairwise ordering. Suffixes are listed in alphabetical order. The first suffix of a pair is listed in the left margin, the second suffix of a pair is shown at the top margin. A '1' denotes attested combinations of the row suffix being followed by the column suffix, a '0' the non-attestation of combinations. *fulA* and *aryA* denote the adjectival suffixes, and *fulN* and *aryN* the nominal suffixes, *lyAJ* denotes adjectival *ly*, and *lyAV* adverbial *ly*.

	age	aryA	aryN	dom	ee	en	er	ery	ess	ette	fold	fulA	fulN	hood	ian	ish	ism	ist	ive	less	ling	lyAJ	lyAV	ment	ness	or	ous	ship	ster	th	wise
age	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0	1	1	0	0	1	
aryA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	
aryN	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	
dom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ee	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
en	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
er	1	0	0	1	0	0	0	1	1	0	0	1	1	0	1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	1	
ery	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
ess	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	1	0	1	
ette	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
fold	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
fulA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	
fulN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
hood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ian	1	1	0	1	0	0	1	1	1	0	0	0	0	1	1	1	1	0	0	1	0	1	0	0	1	1	1	1	0	1	
ish	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	1	
ism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	1	
ist	0	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
ive	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	1	1	0	0	1	
less	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ling	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	
lyAJ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	
lyAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ment	0	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	
ness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	
or	1	0	0	1	0	0	0	1	1	0	0	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	0	0	1	
ous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ship	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	
ster	1	0	0	1	0	0	0	0	0	1	0	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	
th	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	
wise	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	

Figure 1: An example of 15 suffix combinations ruled out by Complexity-based Ordering given 6 attested pairwise suffix combinations. The ordering within the triads *-er*, *-ist*, *-ian* and *-dom*, *-ship*, *-less* is arbitrary. The graph in the right panel summarizes the information in the (partial) adjacency matrix in the left panel. Violations to Complexity-based Ordering would show up as 1s below the diagonal in the adjacency matrix, and as upward pointing arrows in the graph.



Ordering given the 6 attested suffix combinations with *-ess* marked as attested. The right panel shows a directed acyclic graph that is equivalent to the information in the table to its left. Attested combinations are connected by 'edges', with arrowheads specifying the order of the affixes. If a derived word with *-ship* preceding *-ist* were to be discovered, it would introduce a 1 below the diagonal, and an upward pointing arrow in the graph. This upward arrow would introduce a cycle into the graph, and change it from an acyclic into a cyclic graph.

The question that we now have to address is whether Table 3 can indeed be rearranged such that no suffix combinations are left below the diagonal. Or, framed in terms of graph theory, is the graph defined by Table 3 indeed a directed acyclic graph which can be drawn such that all its edges point downwards? Affix pairs for which both orders are attested, such as *-ness* and *-less*, immediately tell us that our graph cannot be acyclic. The question therefore becomes to what extent our graph is surprisingly close to being acyclic.

Figure 2 presents a diagram for our directed suffix graph using the implementation of the algorithm of Gansner et al. (1993) as available in the R (R Development Core Team 2007) package `Rgraphviz` (which is part of the bioconductor project, see <http://www.bioconductor.org>).

In this figure, the overall majority of arrows (151 of 161, shown in grey) point downwards. However, we find 10 upward-pointing arrows, highlighted in black. Seven of these give rise to cycles with two nodes (*aryA-ment*, *ist-aryN*, *ish-ian*, *ist-er*, *ness-less*, *ness-wise*, *ery-ist*)

and three introduce larger cycles (*-ship-ment*, *-ish-ment*, *-lyAJ-ish*). To the best of our knowledge, these 10 exceptions cannot be eliminated, we can at best find a reordering that has again ten exceptions. These 10 exceptions tend to be the same across reorderings, but need not be fully identical. Table 4 presents a rearrangement of the rows and columns of the adjacency matrix Table 3, with the same 10 exceptions appearing below the main diagonal. The small number of exceptions suggests informally that the English suffixes under consideration closely approximate the predictions of the Complexity-based Ordering hypothesis.

At this point, we need to consider how good this approximation is. A problem that arises here is that Figure 2 presents only one of many possible visualizations, and Table 4 represents only one of many possible rearrangements of the adjacency matrix with 10 exceptions below the diagonal. The problem of minimizing the number of upward-pointing arrows (or equivalently, the number of observations below the diagonal in the adjacency matrix), a problem known as the "feedback arc set" problem, is NP-complete, which basically means it cannot be solved in reasonable time (Gansner et al. 1993). To evaluate the extent to which surprise is justified for the observed number of exceptions, we proceeded as follows.

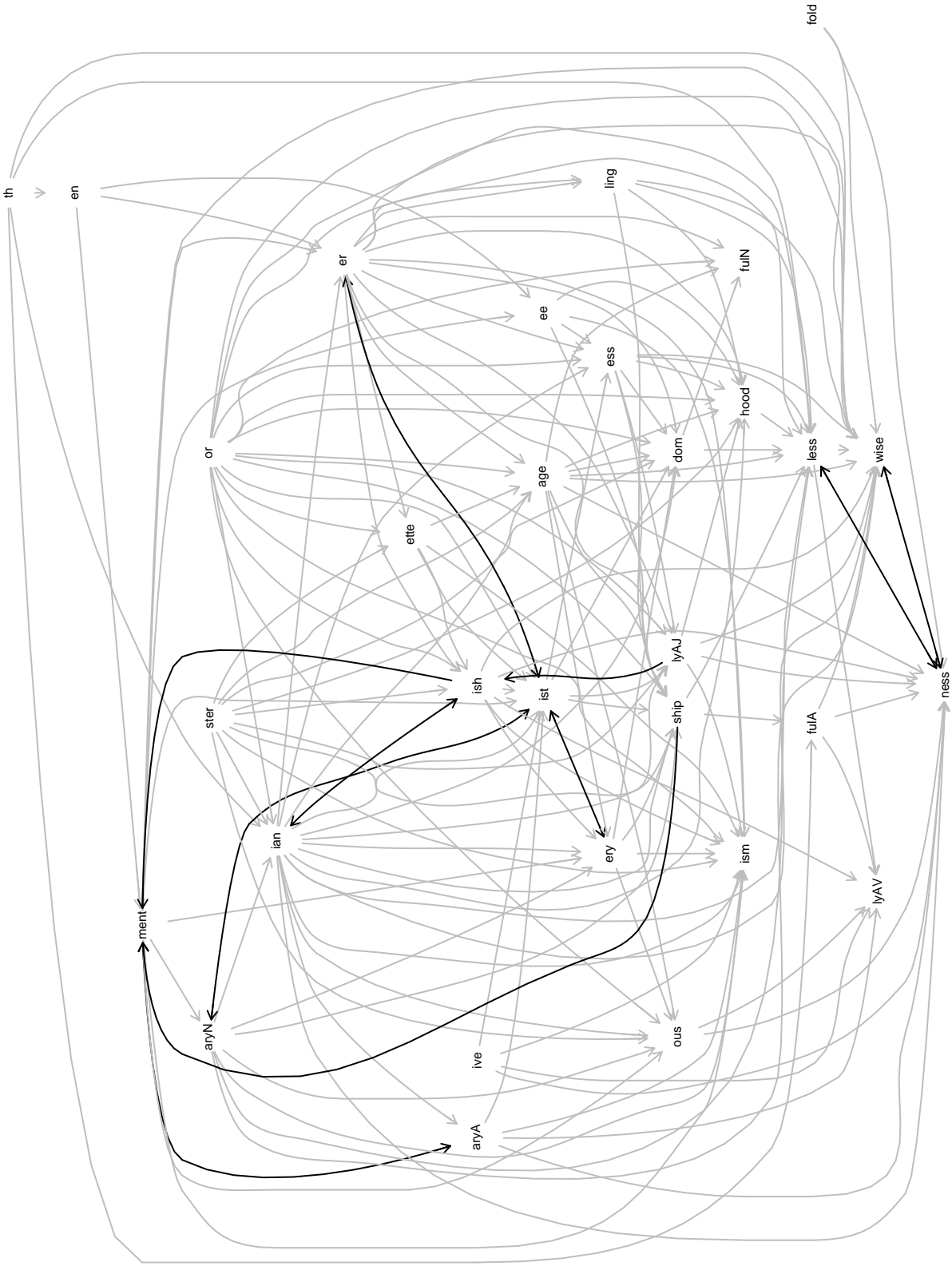
As a first step, we considered the likelihood of observing as few as 10 feedback arcs in a graph with 31 vertices and 161 randomly assigned edges. To estimate this likelihood, we ran 10,000 simulation runs in which 161 edges were randomly assigned, and calculated for each run the number of feedback arcs and loops (edges connecting a vertex to itself). In these simulations, edges were assigned randomly, thereby creating random suffix combinations that are, in general, not attested in English. For the 10,000 runs, the mean number of edges violating acyclicity was 83.15, and the smallest number of such edges (observed once only) was 63. This shows that it is extremely unlikely ($p < 1/10,000$) that the actually observed number of edges violating acyclicity (10) is due to chance.

As a second step, we considered to what extent it is possible to reorder the rows and columns of the simulated adjacency matrices such that as many violations of acyclicity as possible are removed. Using the algorithms described in Gansner et al. (1993) and implemented in the `Rgraphviz` package for the R statistical programming environment, we calculated for each simulation run the number of edges violating acyclicity that could not be eliminated. The mean number of such edges was 70.73, and the minimum was 47. This provides further evidence that the actually observed number of violating edges (10) is much smaller than expected under random conditions.

As a third step, we proceeded to consider the number of violations observed when rows and columns of the empirical adjacency matrix are randomly reordered, while respecting the existing affix combinations. Across 10,000 simulation runs, the mean number of violations was 78.96, with as minimum 40 (observed once only). This provides further confirmation that the observed number of violations (10) is unlikely to be due to chance.

For each of this second set of 10,000 simulated data sets, we also considered to what extent the number of violations might be reduced by application of the algorithm of Gansner et al. (1993). The mean number of exceptions obtained in this way was 21.04, with a minimum of 10 (observed in $21/10,000 = 0.0021$ of the simulation runs). Apparently, the extent to

Figure 2: Directed graph for 31 suffixes. Downward-pointing edges are shown in grey, upward-pointing edges introducing cycles are highlighted in black.



which the algorithm of Gansner et al. is successful depends on the initial order of rows and columns in its input adjacency matrix. Importantly, 10 violations emerged as our best estimate of the minimum number of violations that can be obtained by reordering rows and columns such that the actual affix combinations are respected.

In summary, in an unconstrained, 'parallel-world' scenario, the likelihood of observing only 10 exceptions to acyclicity when 161 edges are assigned randomly in a graph with 31 vertices is vanishingly small, even when exceptions are eliminated to the best extent possible by reordering the suffixes. In a more constrained, 'real-world' scenario in which the real affix combinations of English are inspected and left unchanged in the simulations, the number of violations observed under random reordering of rows and columns remains much larger than 10. In short, under random conditions, the observed number of violations remains extremely unlikely. Finally, reorderings guided by the algorithm of Gansner et al. (1993) show that 10 is our best estimate of the minimum number of violations possible. Table 4 shows one of the suffix rankings for an adjacency matrix with this minimum number of violations.

Let us now consider the exceptional suffix combinations, as listed below the diagonal of Table 4 and represented with upward arrows in Figure 2, in order to assess their theoretical weight. We first checked whether these problematic combinations generate a high number of types by checking their type frequencies in the CELEX lexical database (c. 18 million tokens) and in the written subcorpus of the British National Corpus (c 90 million tokens). Types with these combinations are not attested in either of the two corpora. Table 5 illustrates this point by listing all combinations (with their type frequencies) that can be found in these two sources.⁴ This means that none of the exceptions come from the two corpora, they are all from the OED.

Having observed that the exceptional combinations are extremely rare, we now consider their linguistic properties. We first consider the seven instances where we have double-headed arrows in Figure 2 and where in the adjacency matrix (Table 4) the order below the diagonal has a mirror image above the diagonal, the latter in accordance with Complexity-Based Ordering, the former providing a counterexample. Consider the following pairs, in which the form found below the diagonal is printed in bold.

<i>-aryA</i> and <i>-ment</i> :	militaryment and <i>complementary</i>
<i>-ist</i> and <i>-er</i> :	alchemister and <i>consumerist</i>
<i>-ish</i> and <i>-ian</i> :	Irishian and <i>christianish</i>
<i>-ist</i> and <i>-aryN</i> :	evangelistary and <i>voluntaryist</i>
<i>-ery</i> and <i>-ist</i> :	artillerist and <i>dentistry</i>
<i>-ness</i> and <i>-less</i> :	happinessless and <i>hopelessness</i>
<i>-wise</i> and <i>-ness</i> :	otherwiseness and <i>businesswise</i>

Let us discuss each pair in turn. Of the pair *militaryment* and *complementary*, the first violates Complexity-Based Ordering. Given that the form also violates the selectional restrictions of *-ment*, which takes either bound roots or verbs as bases, *militaryment* can be

⁴Note that we have not attempted to maintain the ranking of Table 4 for Table 5, which was obtained by independent application of the algorithm for minimizing the number of exceptions described above.

Table 5: Adjacency matrix for a Complexity-Based Ordering for 31 English derivational suffixes, using only those suffix combinations that are attested in the CELEX lexical database or in the written subcorpus of the British National Corpus. For each suffix combination, the corresponding number of attested word types is shown.

	th	en	age	er	ster	ette	or	ian	ive	ist	ish	ment	fulN	ery	ous	aryN	less	fulA	lyAJ	ess	lyAV	aryA	ee	ness	hood	ism	wise	ship	fold	ling	dom	
th	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
en	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
age	0	0	1	0	0	0	0	0	1	2	0	0	0	2	0	0	6	5	0	0	0	0	0	0	2	5	21	0	0	0	0	
er	0	0	0	2	0	0	0	0	1	2	0	0	0	0	0	0	0	5	1	0	0	0	0	1	1	0	0	0	0	0	0	
ster	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	
ette	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	
or	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	2	1	5	0	0	0	0	0	0	1	1	0	0	0	0	
ian	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	1	5	2	0	0	1	1	34	3	21	0	0	0	0	
ive	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	91	67	20	0	67	20	0	0	0	0	0	0	
ist	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ish	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	76	80	0	0	80	0	0	0	0	0	0	0	0
ment	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	76	80	0	0	80	0	0	0	0	0	0	0	0
fulN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	98	76	0	0	76	0	0	0	0	0	0	0	0
ery	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	156	76	0	0	76	0	0	0	0	0	0	0	0
aryN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	6	0	0	6	0	0	0	0	0	0	0	
less	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	111	123	0	0	123	0	0	0	0	0	0	0	0
fulA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	1	1	0	0	0	0	0	0	0	
lyAJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	45	1	0	45	1	0	0	0	0	0	0	
ess	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
lyAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	2	0	0	0	0	0	0	
aryA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	
ee	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3	0	0	0	0	
ness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
hood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
ism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
wise	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ship	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
fold	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ling	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
dom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

regarded as truly idiosyncratic. The combination *alchemister*, an earlier form of *alchemist*, has its last citation in the OED in 1586. The (according to the OED) 'nonce-word' *Irishian*, 'an expert in the Irish language' (singularly attested for 1834), is obviously coined in analogy to *Grecian* 'an expert in the Greek language', and not an instance of some regular suffixation pattern. The four remaining cases, *evangelistary* and *voluntaryist*, *artillerist* and *dentistry*, *happinessless* and *hopelessness*, and *otherwiseness* and *businesswise*, are all in accordance with the selectional restrictions of the respective suffixes. These suffix sequences suggest that Complexity-Based Ordering imposes a very strong but nevertheless probabilistic constraint on suffix ordering. The last four exceptional formations all have very small probability in the language, as evidenced by their non-occurrence even in large corpora, but they are possible and available for use.

The remaining three cases, which are represented with upward arrows introducing into the graph cycles with more than two nodes, are *woollyish*, *courtshipment*, and *foolishment*. The latter two forms violate the selectional restrictions of *-ment*, which is a deverbal suffix, and are thus idiosyncratic forms (note that *courtshipment* is singularly attested in 1649, while *foolishment* is a 20th century formation). The derivative *woollyish* is a somewhat controversial case, since it could well be argued that this formation does not involve the adjectival suffix *-ly*, as in *father-ly*, but of adjectival *-y*, as in *hair-y*. The latter analysis is given in the OED, while CELEX provides both possibilities, but it is unclear how one could prove either analysis. We therefore decided to be conservative and treat *woollyish* as a formation involving adjectival *-ly*, and therefore as an exception to strict hierarchical ordering. Under the analysis of the OED, this suffix pair would not even count as a counterexample to complexity-based ordering. In any case, the formation is in complete accordance with the selectional restrictions of *-ly* and *-ish*. In Hay & Plag (2004) the suffixes were in fact ranked in this order, and it is only the attested combinations with suffixes from the new set of suffixes that forced us to rerank *-ish* before *-ly*. The combination as attested in *woollyish*, if indeed taken as an instance of *-ly* and not of *-y*, would provide yet another illustration of the fact that Complexity-Based Ordering is subject to minor leakage when the selectional restrictions offer opportunities for word formation below the diagonal in the adjacency matrix (cf. Table 4). To summarize our discussion of the counterexamples, we can say that there are indeed such exceptions. They are, however, very rare, i.e. they do not occur in the corpora we investigated, but are only attested in the OED. Of these forms, many are very old and probably not in use. Hence, complexity-based ordering emerges as a robust generalization with only little leakage.

The interesting question now is why a suffix hierarchy should exist in the first place. Why should selectional restrictions impose a partial ordering that, with a few exceptions, can be represented as a directed acyclic graph? This is a pure accident under the selectional restriction hypothesis. Hay & Plag (2004) argue that it follows naturally from complexity-based Ordering. According to Complexity-based Ordering, the position of a given suffix in the hierarchy reflects the degree to which that suffix is processed independently of its base. Suffixes higher in the hierarchy, i.e. those that are more to the right in the adjacency matrix, should be more easily separable from their bases than those suffixes that are lower in the

hierarchy, i.e. more to the left in the adjacency matrix. Overall, the most easily separable suffix should be at the right end of the hierarchy, and the least separable suffix at the left end. In order to test whether this is true, one would have to check whether the rank of a given suffix in the hierarchy correlates with the rank of that suffix with regard to independent measures of constituent-driven processing. Recently, Hay and Baayen (2002) have proposed such measures, and we follow Hay and Plag (2004)'s footsteps in employing them to test Complexity-based Ordering.

On the basis of the analysis of 80 English affixes Hay and Baayen show that parsing ratio and are strongly correlated. In general, those affixes that are easily separable from their bases in parsing are also those suffixes that are most productive. This fact is in line with the observation that productive processes are semantically and phonologically transparent. To determine the type and token parsing ratios, Hay and Baayen calculate, for any given affix, in what proportion of words the affix is likely to be parsed, based on the frequency characteristics of the affixes and the words which contain it. Using our example from above again, *-ment* is probably parsed in *discernment* (because *discern* is much more frequent than *discernment*), whereas it is probably not parsed out in *government* (because *government* is more frequent than *govern*). Hay and Baayen also calculate the proportion of tokens containing the affix which are likely to be parsed. The resulting parsing ratios therefore indicate the proportion of types (the type parsing ratio) or tokens (token parsing ratio) containing an affix which are likely to be parsed. For example, if an affix was represented only by words which are unlikely to be parsed, the parsing ratios would be 0. If it was represented only by words which are likely to be parsed, the parsing ratios would be 1. The higher the type (or token) parsing ratio, the greater the proportion of types (or tokens) which are prone to parsing.

For the computation of productivity, Hay and Baayen (2002) used the corpus-based productivity measure \mathcal{P} (Baayen & Renouf 1996), which is the number of hapax legomena, i.e. the words that occur only once in the corpus with a given affix, divided by the number of tokens with that affix. This productivity measure is based on the following reasoning (see, for example, Plag (1999: chapter 5) for more detailed discussion). Assuming that productivity is defined as the possibility of creating a new word, it should in principle be possible to estimate or quantify the probability of the occurrence of newly created words of a given morphological category. By definition, newly coined words have not been used before, they are low frequency words and do not have an entry in our mental lexicon. But how can we understand these new words, if we don't know them? We can understand them in those cases where an available word-formation rule allows us to decompose the word into its constituent morphemes and compute the meaning on the basis of the meaning of the parts. The word-formation rule in the mental lexicon guarantees that even complex words with extremely low frequency can be understood. If, in contrast, words of a morphological category are all highly frequent, these words will tend to be stored in the mental lexicon, and a word-formation pattern will be less readily available for the perception and production of newly coined forms. This means that unproductive morphological categories will be characterized by a preponderance of words with rather high frequencies and by a small number of words

Table 6: Mean CO-ranks for 31 English suffixes, averaged over 4073 orderings with 10 exceptions.

	suffix	rank		suffix	rank		suffix	rank		suffix	rank
1	<i>th</i>	0.89	9	<i>ette</i>	8.62	17	<i>lyAJ</i>	14.58	25	<i>fulA</i>	21.87
2	<i>en</i>	1.78	10	<i>aryA</i>	9.27	18	<i>ery</i>	16.19	26	<i>fold</i>	21.87
3	<i>ment</i>	2.67	11	<i>ive</i>	9.27	19	<i>ling</i>	16.27	27	<i>wise</i>	25.80
4	<i>or</i>	4.40	12	<i>ist</i>	11.44	20	<i>ship</i>	18.63	28	<i>lyAV</i>	25.83
5	<i>ster</i>	4.42	13	<i>ee</i>	11.54	21	<i>dom</i>	18.71	29	<i>fulN</i>	25.86
6	<i>aryN</i>	4.55	14	<i>ish</i>	12.70	22	<i>hood</i>	18.81	30	<i>ism</i>	25.86
7	<i>ian</i>	6.24	15	<i>ess</i>	13.17	23	<i>less</i>	21.78	31	<i>ness</i>	25.88
8	<i>er</i>	7.13	16	<i>age</i>	14.21	24	<i>ous</i>	21.83			

with low frequencies. With regard to productive processes, we expect the opposite, namely large numbers of low frequency words and small numbers of high frequency words.

The crucial point now is that the number of hapax legomena of a given morphological category correlates with the number of neologisms of that category, so that the number of hapaxes divided by the number of tokens of that category can be seen as an estimate of the probability of new formations, and as such as an indicator of productivity (see Baayen and Renouf 1996, Plag 2005, Plag 2003: chapter 3, for further illustration and discussion).

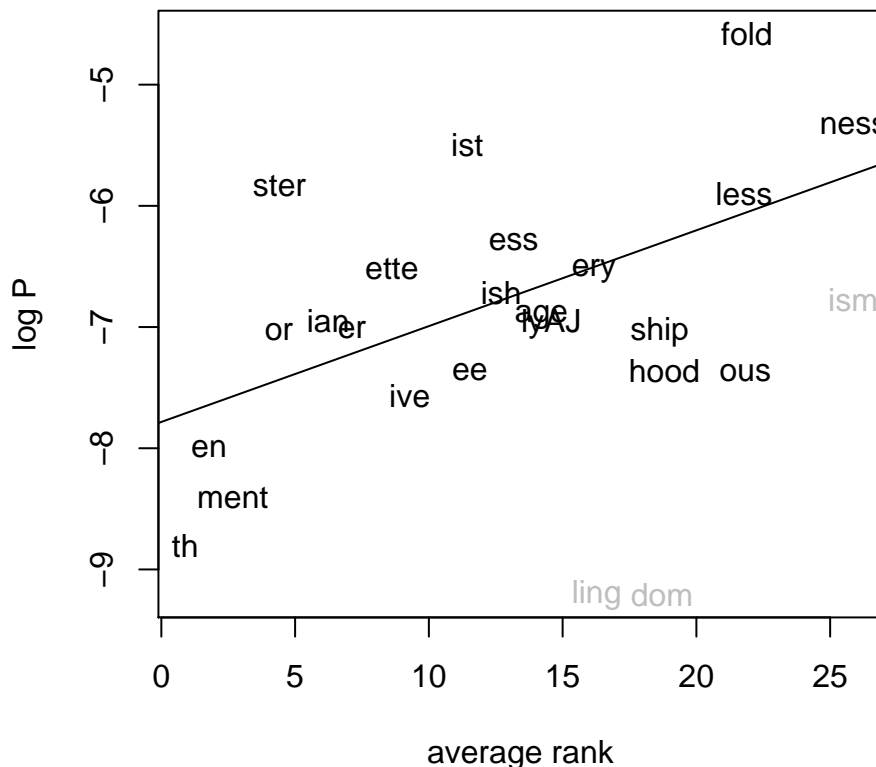
What is important for the present paper is that Hay and Baayen (2002) computed the productivity measure \mathcal{P} with 80 affixes, 25 of which are in the set of suffixes under investigation. We could therefore exploit Hay and Baayen’s results, which were obtained for purposes entirely different from the ones of this paper.

The information from Hay & Baayen (2002) provides us with the possibility to check whether our ‘nearly acyclic’ directed graph can provide us with an affix ranking that is correlated with measures of constituent-driven processing and productivity. The problem that we have to face here is that there are many different orderings with 10 exceptions, and many different diagrams such as shown in Figure 2 can be obtained depending on the specific initial order of rows and columns of the input adjacency matrix. To obtain a ranking that is independent of any of these many particular orderings, we ran a large number of simulation runs which eventually provided us with 4073 adjacency matrices with 10 exceptions. We averaged the ranks of the affixes in these matrices. These averaged ranks are listed in Table 6.

The averaged rank turned out to be correlated only with the log-transformed \mathcal{P} measure, as shown in Figure 3. The regression line (estimated slope 0.079, $t(20) = 3.164$, $p = 0.0049$) was obtained after removal of three influential outliers (shown in grey), as revealed by standard regression diagnostics: the $dfbetas$, the $dffits$, the covariance ratios, Cook’s distance, or the diagonal elements of the hat matrix (cf., for example, Chatterjee et al. 2000). Two of these outliers, *-ling* and *-dom*, have a \mathcal{P} value of 0, which we had changed into 0.0001 before the log-transform.

This means that for the vast majority of suffixes (namely those with a non-zero degree

Figure 3: Productivity as a function of Mean CO Rank. Suffixes shown in grey had a high leverage, and were not included when calculating the regression line.



of productivity), Complexity-based Ordering rank and productivity \mathcal{P} are correlated ($R^2 = 0.33$). The main pattern in the data is that more productive suffixes, i.e. suffixes for which the likelihood that a newly sampled token with that suffix represents an unseen type, tend to have higher mean ranks. Thus we find *-th* and *-en* to the left in the hierarchy with low ranks, and *-ness* and *-less* to the right in the hierarchy with large ranks.

In summary, three facts are noteworthy. First, sequences of English suffixes define a directed graph that is remarkably close to being acyclic. The evidence that this is not due to chance is very strong. Second, there is a substantial amount of indeterminacy in the exact order of the suffixes. Even if the graph would be acyclic, the number of different orderings (called 'topological sorts' in graph theory) can be large. We therefore sampled from the many optimal orders (the orders with a minimum of exceptions) to obtain mean ranks that define an objective linear ordering. Third, the evidence that these mean ranks directly reflect processing complexity is weaker than than for the smaller sample of suffixes in Hay & Plag

(2004). The mean ranks are correlated with only one measure, (\mathcal{P}), instead of three, as in the study by Hay & Plag (2004), and two suffixes with zero productivity \mathcal{P} behave exceptionally. In the next section, we therefore address the question of whether independent experimental evidence can strengthen the potential of Complexity-Based Ordering as a principle of the grammar.

5 Complexity-Based Ordering and Lexical Processing

5.1 The data

Is the rank of a suffix in the ordering hierarchy correlated not only with a distributional productivity measure such as \mathcal{P} , but also with experimental estimates of actual lexical processing costs? And if so, do we expect to find a positive or a negative correlation between Mean CO-Rank and processing cost? If we assume that a suffix that is easily parsed out in comprehension requires less processing time than a suffix that is difficult to parse out, then the prediction follows that suffixes with larger Mean CO-Rank will reveal shorter processing latencies. On the other hand, race models of morphological processing generally assume that look-up in memory is less costly than decompositional processing. Combined with the much larger effect of full-form frequency compared to base frequency observed by Baayen et al. (2008a), this leads us to expect that suffixes with small Mean CO-Rank will enjoy a processing benefit compared to suffixes with large values for Mean CO-Rank.

In order to explore possible answers, we have made use of the behavioral processing measures available in the English Lexicon Project at <http://ellexicon.wustl.edu/> (Balota et al. 2007). This database provides lexical decision reaction times and word-naming latencies for some 40,000 English words.

Lexical decision and word naming tasks have been used extensively by psychologists to study lexical processing of both monomorphemic words (see, e.g. Balota et al. 2004, Baayen et al. 2006) and morphologically complex words (see, e.g. Bertram et al. 2000a,b, Burani & Caramazza 1987, Taft & Forster 1976, Taft 1979, 1988, 2004, Wurm 1997, 2000, Wurm & Aycock 2003, De Vaan et al. 2007). Response latencies elicited in these tasks have afforded considerable insight into a series of factors that have emerged as relevant for understanding lexical processing in language comprehension, although word naming must have an at present not well understood production component as well. In what follows we employ these comprehension data to probe into the balance of computation and storage. In the general discussion we briefly address the question of whether the results generalize to speech production.

To make use of the data from the English Lexicon Project, we compiled a list of all bimorphemic words studied by Hay & Baayen (2002) and extracted all matching entries from the English Lexicon Project website. This resulted in a dataset with 2529 bimorphemic words covering 28 suffixes (data for adjectival *-ary*, adjectival *-ful* and adverbial *-wise* were not available in the English Lexicon Project), with the mean word naming latency for each word and the mean lexical decision latency for each word, averaged over subjects.

To this database we added, for each word, information concerning a wide range of lexical and distributional variables. The properties gauged by these variables range from a word’s form (its length, and the density of its similarity neighborhood) to the complexity of its inflectional (Moscoso del Prado Martín et al. 2004b, Kostić et al. 2003) and derivational paradigms (Schreuder & Baayen 1997, Moscoso del Prado Martín et al. 2004a), and from phonetic properties of the initial segment (which may cause measurement error for the voicekey device registering naming latencies) to frequency measures for the derived word and its base.

Many of the predictor variables that we included in the statistical models reported below are not of interest to the goals of this study when considered in isolation. Nevertheless, we discuss them briefly, for two reasons. First, their inclusion in our statistical models substantially reduces the likelihood that our critical morphological variables are confounded with other hidden processing factors. In other words, we need to make sure that the morphological variables that we are really interested in remain significant predictors in a model that also includes other correlated variables. For instance, word length is negatively correlated with word frequency, hence, frequency effects for derived words and their base words can be assessed properly only once their lengths have been taken into account. Second, it turns out that the joint effect of all variables is important for understanding Complexity-Based Ordering.

We fitted linear mixed effects models (Pinheiro & Bates 2000, Bates & Sarkar 2005b, Bates 2005, Baayen et al. 2008b) to these data, using the `lme4` package (Bates & Sarkar 2005a) in the R statistical programming environment (R Development Core Team 2005), with `Affix` as random effect and a range of fixed-effect lexical predictors. We first consider the word naming latencies, and then the reaction times in visual lexical decision.

5.2 Word naming

Table 7 lists the fixed-effects coefficients⁵ and their associated statistics for all predictors that emerged as significant in a stepwise regression analysis for the word naming latencies.

Inspection of the residuals of our initial model revealed marked departure from normality, which was alleviated substantially by removing data points with outlier residuals (defined as absolute standardized residuals exceeding 2.6). Table 7 lists the estimates of the coefficients for this final trimmed model. We evaluate significance with the help of 10,000 samples from the posterior distributions of the coefficients using Markov chain Monte Carlo (MCMC) sampling (see, e.g., Baayen et al. 2008b). From these samples, we obtained the 95% Highest Posterior Density (HPD) confidence intervals, and the corresponding two-tailed p -values. Since the HPD interval never contains zero, and all p -values are small, we may conclude that all predictors are statistically significant.⁶

⁵The coefficient for a numeric predictor specifies the unit increase (or decrease, if negative) in the (log-transformed) latency corresponding to a unit increase in the value of that predictor, when all other predictors in the model are held constant.

⁶Quadratic terms were included in the lexical decision and naming latency models whenever they improved the fit of the models significantly. Likelihood ratio tests showed significant differences between models

Table 7: Fixed-effect coefficients in a mixed-effects model fitted to the log-transformed word naming latencies of 2529 bimorphemic suffixed English words. **DerFreq**² and **BaseFreq**² denote the quadratic terms for **Derived Frequency** and **Base Frequency**. **Voiced** is the contrast coefficient for voiced as opposed to voiceless initial segments. Estimate: estimated coefficient, HPD lower/higher: 95% Highest Posterior Density (HPD) intervals based on 10,000 Markov chain Monte Carlo samples, p : two-tailed MCMC probability.

	Estimate	HPD lower	HPD upper	p
Intercept	6.5006	6.4351	6.5667	0.0001
A Voiced	-0.0272	-0.0341	-0.0201	0.0001
B Length	0.0096	0.0061	0.0130	0.0001
NSyll	0.0203	0.0127	0.0282	0.0001
C N1	-0.0195	-0.0243	-0.0145	0.0001
H1	0.0041	0.0021	0.0060	0.0001
H3	0.0031	0.0004	0.0057	0.0200
D BNCd	-0.0118	-0.0177	-0.0065	0.0001
E DerFreq	-0.0297	-0.0343	-0.0248	0.0001
DerFreq ²	0.0018	0.0007	0.0030	0.0008
BaseFreq	-0.0068	-0.0094	-0.0042	0.0001
BaseFreq ²	0.0016	0.0008	0.0023	0.0001
SynWord	-0.0131	-0.0228	-0.0035	0.0068
SynBase	-0.0087	-0.0149	-0.0027	0.0068
F LogBigFreq	0.0045	0.0024	0.0068	0.0001
V	0.0303	0.0021	0.0590	0.0390
log(above)	-0.0398	-0.0655	-0.0130	0.0030

Let us go through Table 7 in more detail, starting with section **A**. To control for voicekey artefacts in word naming, we included factors specifying whether the first phoneme was voiced, and whether the first phoneme was a vowel or a consonant. Only the voicing of the first phoneme (**Voiced**) turned out to be a significant predictor for the naming latencies: Voiced initial segments triggered the voice key more quickly and effectively than unvoiced initial segments.

Section **B** of Table 7 lists the effects of two measures for the words' lengths. Longer words elicited longer naming latencies, as expected, both for length evaluated in letters (**Length**)

with the quadratic terms as against ones without these terms (**Base Frequency** quadratic term: $X^2_{(1)} = 4.293, p = 0.0383$ for lexical decision, $X^2_{(1)} = 8.417, p = 0.0037$ for the naming latency; **Derived Frequency**: $X^2_{(1)} = 4.035, p = 0.0446$ for lexical decision; **Length**: $X^2_{(1)} = 13.75, p = 0.0002$ for lexical decision. It is only the quadratic term of **Derived Frequency** in naming latency that receives less support from the likelihood ratio test ($X^2_{(1)} = 3.036, p = 0.0815$). In this particular case we decided to nevertheless leave the term in the model, due to its being both highly significant in the model and of prime theoretical interest. We add the likelihood statistics for completeness only and refer the reader to (Pinheiro & Bates 2000:chapter 2) for critical discussion of the use of likelihood ration tests for comparing models wiht different fixed effects. These authors recommend the use of conditional tests as documented in our tables 7 and 8.

and for length in syllables (*NSy11*).

Section **C** reports variables gauging the role of orthographic neighbors. A word's neighbors are often defined as those words that differ in one letter or phoneme (*house/mouse*). The total number of such neighbors is often considered as a measure of lexical competition in visual and auditory comprehension (Coltheart et al. 1977, Balota et al. 2004, Luce 1985, Pisoni et al. 1985) and speech production (Scarborough 2004, Vitevitch 2002, Vitevitch & Stamer 2006, Wright 2004). For the present data we did not observe significant effects of the total count of orthographic neighbors (for a similar result for monomorphemic words see Baayen et al. 2006). However, *N1*, the count of neighbors of the base (evaluated for lemmas in CELEX) that differ only with respect to the initial phoneme (e.g., *deal* and *veal*) was predictive: The greater the number of neighbors sharing the remainder of the base word, the shorter the naming latency was. Other measures effectively gauging phonological co-activation are Shannon entropies calculated for the cohort competitors at the first (H_1) and the third (H_3) segment of the word (evaluated against the wordforms in the CELEX lexical database). These entropy measures indicate that the target entered into a process of competition with words sharing the words' initial segments (see Baayen (2007) for further discussion of these measures).

The predictor labeled *BNCd* represents the frequency of the derived word in the demographic subcorpus of the British National Corpus. This subcorpus samples spontaneous spoken English. We included this variable as a correction on the written frequencies (obtained from CELEX), particularly because it captures to a considerable extent differences in age of acquisition (Baayen 2005, Baayen et al. 2006). By bringing *BNCd* into the model, we reduce the risk of confusing effects of frequency of exposure with age of acquisition, a variable that by itself may have more explanatory power than word frequency (Brybaert et al. 2000, Lewis et al. 2001, Morrison & Ellis 2000) but that generally is not available for complex words.

The variables in sections **A** to **D** of Table 7 were included, as mentioned above, to bring under statistical control a range of non-morphological variables that are known to play a role during lexical processing. Sections **E** and **F** of Table 7 bring us to the morphological variables that are of primary interest to us here. Section **E** discusses measures relating to properties of the derived words and their base words, section **F** discusses measures representing properties of the suffix.

Section **E** begins with listing the coefficients for the frequency of the derived word (collapsing inflectional variants), denoted by *DerFreq*, and the frequency of its base word (again collapsing over inflectional variants), denoted by *BaseFreq*. These two frequency measures have often been interpreted as measures of memory-based and rule-driven processing (see, e.g. Pinker 1991, 1997, 1999, Baayen et al. 1997). However, it is more likely that *Derived Frequency* taps into procedural memory traces for complex words (i.e., past experience with parsing and producing the complex word), and that *Base Frequency* taps into the general availability of the base word (see Baayen 2007, Balling & Baayen 2007, Taft 2004, Baayen et al. 2008a).

We modeled the nonlinear effects of *Derived Frequency* and *Base Frequency* by means

of quadratic polynomials.⁷ The functional relation between these frequency measures and naming latency is now represented by part of a parabola, instead of by a line segment. The mathematical equation for a parabola is $y = a + bx + cx^2$, where a is the intercept, b is the *linear coefficient*, and c is the *quadratic coefficient*. Straight lines can be viewed as a special case with $c = 0$, in which case b is the coefficient for the slope of the line. In Table 7 the linear coefficients of the two polynomials are referenced as **DerFreq** and **BaseFreq** and the corresponding quadratic coefficients as **DerFreq²** and **BaseFreq²**. As the individual coefficients of a parabola are somewhat less straightforwardly interpretable, their joint effect is visualized in the two upper panels of Figure 4. (The lower panels show the corresponding effects in visual lexical decision, to be discussed below.)

What Figure 4 tells us is that in word naming the facilitatory effects of frequency level off for higher frequencies, and more so for **Base Frequency** than for **Derived Frequency**. In fact, the facilitation obtained from **Base Frequency** is restricted to two thirds of the frequency range. We return to discuss these frequency effects and their theoretical consequences for relative frequency (Hay 2001) and Complexity-Based Ordering below. Here we restrict ourselves to pointing out that the frequency of the derived word is a much stronger predictor than the frequency of the base word. A change of one log **Derived Frequency** unit in the left panel corresponds to a much larger reduction in log naming latency than a unit change in **Base Frequency**. This differential effect is indicated numerically by the difference between the two linear coefficients in Table 7 (-0.0297 for derived frequency vs. -0.0068 for Base Frequency). This highlights the importance of past experience with complex words.

The last two measures in section **E** of Table 7 gauge the role of the number of meanings carried by the derived word and by its base. In a fully decompositional system with storage being reserved to irregular forms, the set of words closely related semantically to the base should be far more important for predicting lexical processing than the set of words related semantically to the derived words. The more the balance of storage and computation shifts towards storage, the greater we expect the relative contribution of the set of words related to the derived word itself to be.

We estimated numbers of meanings by means of the synonym sets (synsets) in WordNet (Miller 1990, Fellbaum 1998). For each derived word and each base word, we counted the number of synsets that they are listed in (**SynWord** and **SynBase** respectively). For both counts, we observe facilitation. This finding contrasts with the results obtained by Baayen et al. (2006) for monomorphemic monosyllabic words in English. For such words, the synset measures were predictive for visual lexical decision, but not for word naming. The present results show that when morphologically complex words have to be named, derivatives that themselves have denser semantic networks, and also derivatives with base words with denser networks have a small advantage in word naming as well.

Two explanations suggest themselves for this difference. Complex words might lead to deeper semantic processing than do simplex words. However, a linguistically less interesting

⁷See Balota et al. (2004) and Baayen et al. (2006) for nonlinear functions relating word frequency and response latencies for monomorphemic monosyllabic words, and Harrell (2001) for in-depth discussion of the modeling of nonlinear functions in multiple regression.

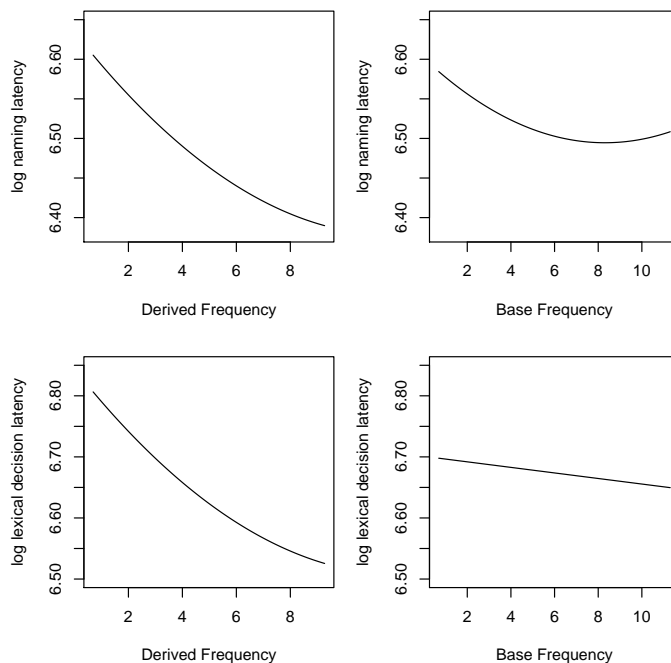


Figure 4: Partial effects of **Derived Frequency** (left) and **Base Frequency** (right) for the log-transformed word naming latencies (upper panels) and visual lexical decision latencies (lower panels) of suffixed derived words.

explanation calls attention to a difference in the make-up of the experimental lists used in the English Lexicon Project. For monomorphemic words, participants were exposed to monomorphemic words only. This may have led to shallow semantic processing. For the derived words, however, participants were exposed also to many complex words. This may have favored deeper semantic processing. Further research is required to evaluate the merits of these two alternatives. Interestingly, even though the range of number of synsets is smaller for the derived word than for its base, the facilitatory effect for the derived count is larger (cf. the coefficients of -0.013 for **SynWord** as against -0.009 for **SynBase**). This is consistent with the greater effect of **Derived Frequency** compared to **Base Frequency**, and supports the hypothesis that memory traces for complex words play an important role in lexical processing.

The final section of Table 7 turns to measures relating to the suffix. The predictor labeled **LogBigFreq** denotes the logarithmic transform of the orthographic bigram spanning the boundary between base word and suffix (e.g., **df** in **handful**). A greater frequency of the morphologically critical bigram correlates with longer naming latencies. A higher bigram frequency has been noted to render a word less decomposable (e.g., Seidenberg 1987, Andrews 1992, Hay 2003). Here we replicate the finding that decreased decomposability leads to delayed processing.

Surprisingly, affix productivity, as gauged by the number of types as listed in Hay &

Baayen (2002) for bimorphemic derived words in English, here denoted by **V**, is inhibitory for word naming. A possible explanation for this unexpected inhibition is that a productive suffix might activate a larger range of base words, and that these base words subsequently compete with the base that is to be named. If this explanation contains a grain of truth, we should find that in lexical decision, a task that does not require the selection of a base for articulation, no inhibitory effect of **V** should be present. We return to this issue below.

The very last predictor in Table 7, labelled $\log(\text{above})$, is the log of the number of types above the parsing line, one of the productivity measures mentioned previously that were introduced by Hay & Baayen (2002). Somewhat to our surprise, given the crude way in which Hay and Baayen estimated the parsing line — on the basis of a single suffix (*-ness*) only, using a very simple computational model — the number of types above the parsing line $\log(\text{above})$ is a solidly significant predictor of naming latencies. Exactly as expected, a greater number of types above the parsing line correlates with decreased naming latencies. Other things being equal, suffixes that occur in large numbers of derivatives that have a high likelihood of being parsed, themselves afford faster lexical processing — at least as gauged by word naming — for *any* derivative in which they occur. If this result generalizes to visual lexical decision, the present evidence that parsing speeds lexical processing would be strengthened.

We conclude our analysis of the naming data by discussing the random effects structure of our model. In our analyses, **Suffix**, a factor with 27 levels, was incorporated as a random effect since there are many suffixes that we have not included in our sample. As a consequence, the coefficients of our models listed in Table 7 allow us to predict processing costs for the average unseen suffix. For the suffixes that actually appear in our sample more precise predictions can be made, however, by fine-tuning these coefficients to each suffix individually.

In mixed-effects modeling, such fine-tuning is always required for the intercept.⁸ As the intercept is a kind of grand average that expresses the baseline latency required for the task, fine-tuning the intercept for our data therefore amounts to allowing baseline processing costs to vary from suffix to suffix. The estimated standard deviation for the random intercepts for **Affix** was 0.0265.

For the present data, a likelihood ratio test shows that further fine-tuning is required for the linear coefficient of **Derived Frequency** ($p < 0.0001$). The facilitatory linear effect of **Derived Frequency** varied slightly but significantly from suffix to suffix. The standard deviation estimated for the random slopes for **Derived Frequency** was 0.0079. A parameter for a correlation of the random slopes and intercepts proved to be superfluous.⁹ We complete the description of our model with the estimate for the standard deviation for the residual error, which was 0.0797, and turn to the lexical decision data.

⁸For details on random effects in mixed-effects models, the reader is referred to Pinheiro & Bates (2000), Faraway (2006), and for nontechnical introductions to Baayen (2008) and Baayen et al. (2008b).

⁹In order to avoid spurious correlations in the random effects structure of the model, **Derived Frequency**, **Base Frequency** and **Length** were centered, see, e.g., Pinheiro & Bates (2000:30–37) and Baayen (2007:276–277).

Table 8: Fixed-effect coefficients in a mixed-effects model fitted to the log-transformed visual lexical decision latencies of 2529 bimorphemic suffixed English words. **DerFreq**² and **BaseFreq**² denote the quadratic terms for **Derived Frequency** and **Base Frequency**. Estimate: estimated coefficient, HPDlower/higher: 95% Highest Posterior Density (HPD) intervals based on 10,000 Markov chain Monte Carlo samples, *p*: two-tailed MCMC probability. MCMC estimates are based on 10,000 samples. Variables are grouped as in Table 7.

		Estimate	HPD lower	HPD upper	<i>p</i>
	Intercept	6.6721	6.6090	6.7346	0.0001
B	Length	0.0159	0.0103	0.0219	0.0001
	Length ²	0.0031	0.0021	0.0041	0.0001
	NSyll	0.0187	0.0097	0.0270	0.0002
C	N1	-0.0094	-0.0155	-0.0035	0.0016
D	BNCd	-0.0196	-0.0261	-0.0132	0.0001
E	BaseFreq	-0.0045	-0.0075	-0.0015	0.0026
	DerFreq	-0.0385	-0.0431	-0.0336	0.0001
	DerFreq ²	0.0023	0.0010	0.0036	0.0016
	SynWord	-0.0280	-0.0391	-0.0165	0.0001
	SynBase	-0.0126	-0.0195	-0.0054	0.0001
F	LogBigFreq	0.0040	0.0014	0.0066	0.0032
	log(above)	-0.0157	-0.0273	-0.0034	0.0108

5.3 Visual lexical decision

The fixed-effects coefficients and their statistics in the mixed-effects model fitted to the visual lexical decision latencies are summarized in Table 8. This model was obtained using the same analytical procedures as for the naming latency, and the general pattern of results is fairly similar. There are four relatively minor differences.

First, **Length** (section **B** of Table 8) shows a U-shaped relation with reaction time, with initial facilitation followed by inhibition, instead of the straightforward linear inhibition observed for word naming. This means that very short words are slower to recognize than medium-length words, but beyond a certain point, additional length makes words slower to recognize. As for the frequency variables in word naming, we modeled this nonlinearity by means of a quadratic polynomial. The linear term is denoted by **Length** in Table 8, and the quadratic term by **Length**². The present nonlinear relation between word length and decision latency replicates the results reported by New et al. (2006).

Second, the positional entropy measures H_1 and H_3 were not predictive for lexical decision (section **C** of Table 8). This supports the hypothesis of Baayen (2007) that these measures specifically gauge cohort-like competition during phonological articulation (see also Van Son & Pols 2003, Van Son & Van Santen 2005).

Third, the effect of **Base Frequency** (section **E** of Table 8) is smaller compared to the effect of **Derived Frequency**, and linear rather than nonlinear, as shown in the lower panels

of Figure 4.

Fourth, recall that we observed an inhibitory effect in word naming of V , the number of bimorphemic words in which the suffix occurs. We hypothesized that a productive suffix that occurs in many other derived words co-activates these derived words, thereby creating additional competition for the base of the derived word that is to be articulated. In visual lexical decision, the V measure again had a positive coefficient, but it failed to reach significance ($p > 0.3$). Since visual lexical decisions can be based on global lexical activation and do not require selecting a specific base for articulation, the absence of inhibition in this task is in line with our explanation for the naming data. Possibly, the null effect of V in visual lexical decision is due to the increased competition for the base being cancelled by the advantage of increased general lexical activation, which in general affords faster lexical decision latencies.

Finally, we note that the log-transformed count of types above the parsing line, $\log(\text{above})$, was significant (section **F** of Table 8), just as in word naming.

We complete the summary of the mixed-effects model for the visual lexical decision latencies with the random effects parameters: a standard deviation of 0.0294 for the random intercepts, a standard deviation of 0.0062 for random slopes for **Derived Frequency**, a standard deviation of 0.0068 for random slopes for **Length**, and a standard deviation of 0.0932 for the residual error. Both random slopes were supported by likelihood ratio tests (all $p < 0.02$). Parameters for correlations between random intercepts and random slopes did not have explanatory value and were therefore not incorporated into the model.

Having completed the detailed description of the model for the lexical decision data we now summarize the differences and similarities between the two. Overall, the differences are minor. In lexical decision we find no entropy effects, the base-frequency effect is linear instead of non-linear, and there is no effect of type frequency (and, trivially, no voice key effect). In the naming latencies the effect of length was linear rather than non-linear. The commonalities outweigh these differences by far. In both tasks we see effects of length, of the number of syllables, of lexical neighbors, of base frequency, of derived frequency, of the number of synonyms, of the bigram frequencies, and of the number of formations above the parsing line. Most important for the present discussion is the absence of an effect of the rank in the hierarchy in both tasks, which means that this rank is not predictive for the costs of processing of individual words in comprehension.

We note in addition that there is no role for a straightforward measure of *relative* frequency in these models. Relative frequency was never predictive in our models. Relative frequency is defined simply as the ratio of base frequency to derived frequency. However, in our models, the two frequency measures have very different coefficients, i.e., their own

regression weights, and also different functional shapes (linear vs. non-linear).¹⁰

Furthermore, the random effects structure of our models show that the item-specific balance of rote and rule is modulated by the suffix-specific coefficients for **Intercept** and **Derived Frequency** in both tasks. We conclude that the base frequency and derived frequency are both highly relevant, but that their relative contributions at the item level are more intricate than can be captured by a simple ratio, at least as gauged by the visual lexical decision and word naming tasks.

5.4 Mean CO-Rank and average by-suffix processing costs

In the analyses described above, **Mean CO-Rank** never emerged as a significant predictor variable, i.e. it failed to be predictive for the response latencies at the fine-grained level of the individual items in word naming and visual lexical decision. However, it turns out that this **Mean CO-Rank** is a relevant predictor for processing complexity at a higher aggregation level, that of the suffix itself, i.e. when we average the latencies over all items with a given suffix. This predictivity emerges from an inspection of the median processing latencies for the suffixes. We opted for the median rather than the mean because distributions of reaction times tend to be skewed. For skewed distributions, the median is a better characteristic of the most typical value than the mean. We calculated the median latencies on the basis of the fitted (log-transformed) latencies, which provide our best theoretically informed estimates of suffixal processing costs. By using the median we removed the by-observation noise, and by using the fitted latencies (instead of the observed ones) we factored in all the processing factors discussed in the previous two subsections.

Apart from **Mean CO-Rank** as predictor variable we included in our analysis two other predictor variables that are of potential interest. First, we wanted to know whether potential results would be robust across the two kinds of experiments and we therefore pooled the data from the naming and lexical decision experiments, and included **Experiment** as a predictor, with 'Lexical Decision' and 'Naming' as factor levels. Furthermore, we tested for potential effects of **Stratum** ('Latinate' versus 'Germanic') in order to rule out the possibility that the curvature in the graphs of Figure 5 is due exclusively to Latinate affixes simply having higher processing costs than Germanic affixes, due to, for instance, more complex consonant clusters or later age of acquisition. The suffixes *-age*, *-ary*, *-ette*, *-ian*, *-ism*, *-ist*, *-ive*, *-ment*, *-or*, *-ous* were classified as Latinate, and the suffixes *-dom*, *-ee*, *-en*, *-er*, *-ery*, *-ess*, *-fold*, *-ful*,

¹⁰More formally, note that relative frequency enters into a regression model as a single term that receives a single coefficient. As frequency effects are logarithmic in nature it makes sense to consider relative frequency on a logarithmic scale as well, i.e. as a log odds ratio. Since this log odds ratio is modeled with a single coefficient, say β , and denoting the two frequencies by f_1 and f_2 respectively, we have that

$$\beta \log \frac{f_1}{f_2} = \beta \log f_1 - \beta \log f_2.$$

Note that in this equation the two frequencies receive exactly the same beta weight, and that both are conceptualized as linear effects. This contrasts with the empirical data in which the two frequencies have very different weights and may even differ in their functional form.

Table 9: Coefficients for the mixed-effects model fitted to the median by-suffix latencies. For **Stratum**, the reference level was 'Germanic', for **Experiment**, the reference level was 'Lexical Decision'. **Mean CO-Rank²** denotes the quadratic term for **Mean CO Rank**. Estimate: estimated coefficient, HPDlower/higher: 95% Highest Posterior Density (HPD) intervals based on 10,000 Markov chain Monte Carlo samples, p : two-tailed MCMC probability.

	Estimate	HPD lower	HPD upper	p
Intercept	6.5756	6.5274	6.6247	0.0001
Stratum=Latinate	0.0777	0.0385	0.1160	0.0002
Experiment=Naming	-0.0962	-0.1058	-0.0859	0.0001
Mean CO-Rank	0.0115	0.0044	0.0188	0.0042
Mean CO-Rank ²	-0.0003	-0.0006	-0.0001	0.0088
Stratum=Latinate:Experiment=Naming	0.0193	0.0024	0.0365	0.0310

-hood, -ish, -less, -ling, -ly, -ness, -ship, -ster, -th as Germanic. The classification of affixes into one or the other stratum is sometimes controversial, see, for example, the discussion in Giegerich (1999). We have used the standard criteria as summarized in section 2, i.e. etymology, possible attachment to bound roots, phonological integration, and productivity, ending up with a classification similar, though not identical, to that of previous authors (e.g. Spencer, 1991:79; Hay 2002).

Based on these considerations, we fitted a linear mixed-effects model to the joint data from the naming and lexical decision experiments, with **Suffix** as random effect and **Mean CO-Rank**, **Stratum** and **Experiment** as fixed effects. Table 9 summarizes the coefficients of the fixed effects in the resulting model.

Figure 5 visualizes the relation between median latency and **CO-Rank** for word naming (left panel) and visual lexical decision (right panel). Each panel shows two inverse-U shaped regression curves. The upper curves represent the Latinate suffixes, the lower curves the Germanic suffixes.

In Figure 5 Latinate suffixes are shown in upper case letters. Among these, *-age* is atypical in that its median latency is at least 100 ms shorter than the median latency of any other Latinate suffix. We have no explanation for why *-age* shows this exceptional behavior. As it exerted undue leverage in our statistical models, we removed it from our data set. The regression lines shown in Figure 5 as well as Table 9 are all based on the data with *-age* excluded, and illustrate the main patterns in our data.

The main effect of **Stratum** is as expected: words with Latinate suffixes elicited longer latencies. The interaction of **Stratum** by **Experiment** indicates that this effect was more pronounced for the naming latencies, probably due to increased phonological complexity leading to delays in articulation.

Finally, we observe a non-linear effect of **Mean CO-Rank** that was the same across experiments and strata. In order to check that the effect of **Mean CO-Rank** is not confounded with the productivity measure \mathcal{P} (with which it is correlated), nor with the measure quantifying the number of types parsed (which was predictive at the level of individual words), we in-

retrieval. But parsing is a complex operation that also comes with its own disadvantages. As we move to the left in the Complexity Ordering, the complexity of this operation increases due to phonotactic and frequential properties becoming increasingly similar to those of simplex words.

We can make the opposing forces of rule and rote more explicit with the help of a mathematical model that is related but not identical to the mixed-effects model reported in Table 9. This regression model approximated the inverse U-shaped relation between **Mean CO-Rank** and median by-affix latency in Figure 5 by means of a quadratic polynomial. Expressed in the standard mathematical form of an intercept, a linear term, and a quadratic term, with x representing **Mean CO-Rank**, we have for median latency y that

$$(1) \quad y = a + bx + cx^2.$$

There are many other functional forms that might be considered and that might provide a good fit to the data. In order to highlight the opposing forces of memory and computation, we consider a function that is very similar to the polynomial in 1, with two changes. First, we constrain the coefficients b and c to have the same absolute value. Second, we rescale the predictor x to an interval in $(0, 1)$ representing the balance of storage and computation. At $x = 0.5$, storage and computation are fully balanced. The equation in 2 for median latency y as a function of the storage-computation parameter x' ,

$$(2) \quad y = a' + c' \cdot \underbrace{(1 - x')}_{\text{computation}} \cdot \underbrace{(x')}_{\text{storage}}, \quad (c' > 0),$$

highlights that increasing x' to favor computation ($1 - x'$ smaller, so shorter latencies y) goes hand in hand with a commensurate storage penalty (x' greater, so larger latencies y), and that decreasing x' to favor memory-driven processing likewise is accompanied by a cost in parsing efficiency. For $x' = 0.5$, where the two forces are balanced, processing costs are maximal.

In order to fit model in 2 to the data, we transformed the ranks x as given in Table 6 to values x' in a subinterval (r_{\min}, r_{\max}) in $(0, 1)$. This subinterval was determined by means of a grid search that optimized the R^2 for a linear mixed-effects model with as predictors **Experiment**, **Stratum**, the interaction of these two predictors, and $(1 - x') * x'$ as predictor for Complexity-Based Ordering. The minimum value (r_{\min}) for the transformed rank was estimated at 0.2889, and the maximum value (r_{\max}) at 0.5556. The estimate of the coefficient c' for the transformed CO-Rank was 1.83 ($p < 0.002$ across 1000 Markov chain Monte Carlo samples of the posterior distribution of the parameters). The log likelihood of this new model (108.6) was slightly higher than the log likelihood of the model with the quadratic polynomial (94.9). We note, however, that this small advantage in the log likelihood comes with the cost of an extra parameter: instead of a linear and a quadratic term for x we now have a single parameter c' for the term $x'(1 - x')$, but r_{\min} and r_{\max} add two further parameters. For the present purposes, we regard the original model and the new model with the transformed ranks as equivalent. We call attention to the model with transformed ranks because this model facilitates understanding the pattern of results in terms of a balance between storage and computation.

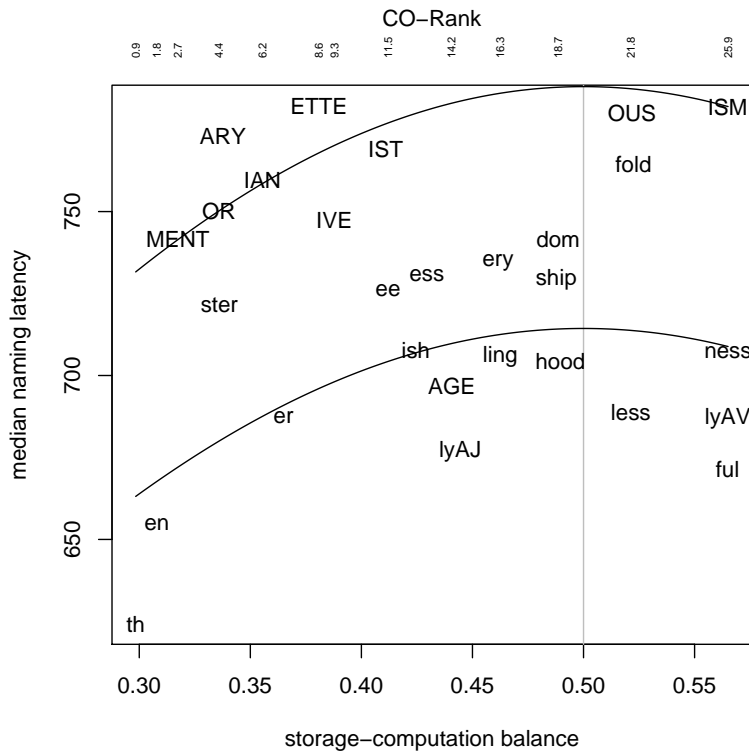


Figure 6: By-suffix median word naming latencies as a function of the balance of storage and computation.

Figure 6 shows the fit of model 2 to the median naming latencies, which are shown on the vertical axis. (The pattern for the lexical decision latencies is very similar — recall Figure 5 — and is not repeated here.) The original Mean CO-Ranks are shown at the top of the graph, the transformed ranks are shown at the bottom. It is easy to see that the maximum naming latency is reached when the balance parameter x' equals 0.5 (highlighted by the grey vertical line). When the balance parameter is increased beyond 0.5, favoring computation, the more productive suffixes receive a processing advantage. When it is decreased, favoring storage, the less productive suffixes receive a processing advantage. The asymmetrical location of the maximum is noteworthy: Overall, memory-based processing apparently offers greater advantages than decomposition-driven processing.

Two final remarks are in order. First, the nonlinearity visualized in Figure 6 depends crucially on the adjacency matrix being as complete as possible. When using only the data from CELEX and the BNC, the nonlinearity is no longer significant, and the relation between Mean Rank and average latency is strictly linear, with positive slope. Apparently, we need to take into account the full range of possibilities offered by English morphology in order to be able to discern the — modest — processing advantage offered by parsing.

Second, we note that for the full data set (see Table 3) Base Valency (the number of

non-zero entries in a suffix' row in the adjacency matrix, i.e., the number of suffixes that can follow it) and Derived Valency (the number of non-zero entries in its column, i.e., the number of suffixes that it can follow) add up to a constant, modulo by-observation noise. This follows from an ordinary least squares regression model in which Base Valency is regressed on Derived Valency :

$$(3) \quad \text{Base Valency} + 0.38 * \text{Derived Valency} = 7.17$$

($F(1, 29) = 4.59, p = 0.041$). In other words, a consequence of Complexity-Based Ordering, given an adjacency matrix that is not too sparse, is that less productive suffixes are more productive as input for further word formation. This result was previously obtained, albeit in a completely different way, by Krott et al. (1999).

From this perspective, the claim advanced by Hay (2002:527–528) that *an affix which can be easily parsed out should not occur inside an affix which can not* can be made more precise. A priori, it is unclear why a more parsable suffix should not be free to occur first or second. Any difficulty in parsing has to be met at some point in time, and why not earlier than later? To the extent of our knowledge, there are no proposals for syntax that it would be advantageous for processing if parsing difficulties are encountered earlier rather than later in the sentence. The crucial point is that morphological complexity is like a coral reef, with upper layers building upon preexisting lower layers which depend more and more on memory, and for which parsing becomes more and more costly.

6 General Discussion

Hay & Plag (2004) addressed the hypothesis of Complexity-Based Ordering developed by Hay (2002, 2003). They showed, for a sample of 15 suffixes, that a partial ordering of these suffixes could be established such that the rank of a suffix in this partial ordering correlated with its productivity, as gauged by measures assessing processing complexity. They also showed that selectional restrictions were in accordance with the complexity-based ranking, and tightened this ranking by providing further constraints on affix ordering.

The present study extends their research in several ways. First, we have broadened the empirical basis for English by extending the ordering to 31 suffixes. This extended set of suffixes showed the same kind of hierarchical behavior, with the ranks in the hierarchy correlating with the respective suffixes' productivity \mathcal{P} .

Second, mixed-effects models fitted to word naming latencies and visual lexical decision reaction times documented significant effects for suffix-related distributional measures, including a productivity measure introduced by Hay & Baayen (2002), the number of types above the parsing line. However, at the level of the individual word and its processing latency, Complexity-Based Rank failed to reach significance. This suggests that it is unlikely that Complexity-Based Ordering should be interpreted as a causal factor in lexical processing, i.e. in the processing of individual words.

Third, however, we were able to show for both word naming and visual lexical decision that at the aggregate level of by-suffix median processing costs, the Mean CO-Rank is

predictive. Interestingly, suffixes with *extreme* ranks in the Complexity Ordering hierarchy enjoy the lower processing costs. We have shown how the inverse U-shaped effect of Mean CO-Rank can be understood as resulting from the opposing forces of storage and computation. As we move down the hierarchy to the lowest values of Mean CO-Rank, complex words become more like simplex words, and retrieval from memory becomes less and less costly. As we move up the hierarchy, junctural and frequential properties allow parsing to operate more effectively, and for the suffixes with extreme ranks, the advantages of parsing become visible in the form of a negative relation between Mean CO-Rank and median processing latency.

The intermediately ranked suffixes have the greatest processing costs. They do not enjoy the full advantages of parsing, nor the full advantages of storage. This situation may have further adverse effects. Suppose, for ease of exposition, that the two access routes, rule and rote, are roughly equally fast for intermediately ranked suffixes. Given that derived words often carry idiosyncratic shades of meaning, and that by definition rule-based processing will lead to fully compositional meanings, intermediately ranked suffixes are more likely to suffer from the semantic ambiguity caused by the simultaneous availability of the opaque and transparent readings (cf. Schreuder et al. 2003). This might lead to delays in lexical processing. Another complication may be the well-known tension between phonological and morphological optimality. In the spirit of Burzio (2002), we note that the words with properties that make them eminently parsable thanks to very low transitional bigram frequencies may also be the words that are more difficult to pronounce due to the presence of more complex consonant clusters. In the light of these considerations, we believe that the balance model introduced in (2) represents multiple

constraints simultaneously: Values of the balance parameter greater than 0.5 bear witness to enhanced constituent-driven processing and greater semantic transparency, values less than 0.5 testify to less marked phonological structure, increased semantic idiosyncrasy, and greater accessibility in lexical memory for phonological sequences.

We have seen that there are more suffixes with a storage-computation balance parameter less than 0.5 than suffixes with a balance greater than 0.5. This suggests that storage in memory has an a-priori advantage over decompositional computation. This may help explain the general trend for affixes to gravitate towards the unproductive over time. Combined with some principle of least effort (Zipf 1949), we can predict that processing load tends to be reduced by simply remembering complex derived words rather than reconstructing them from first principles (using word-formation rules, analogy, or other conceivable mechanisms).

We also note that the asymmetrical location of the maximum of the storage-computation curve in Figure 5 does not come as a surprise in exemplar-based approaches to morphological processing (Bybee 2001, Baayen 2007). Morphological rules, when viewed as generalizations (in lexical memory) over exemplars (in lexical memory), exist thanks to the availability of storage, so it is only natural that storage should have a head start vis-a-vis computation.

One question raised earlier in this paper (see section 5.1) is whether our results, which were based on comprehension data, generalize to speech production. In Levelt's model of speech production (Levelt et al. 1999), which is still the most comprehensive and most widely

accepted model of speech production to date, is fully decompositional in nature and whole-word access to morphologically complex words is ruled out a priori, and productivity does not play a role. Viewed from this perspective, storage-computation balance is a non-issue. However, we have to note that Hay's arguments for graded decomposability and Complexity-based Ordering are based in part on evidence obtained from the acoustic analysis of what speakers actually say. For example, Hay (2001) shows that the stem-final [t] in words such as *swiftly* (decomposition bias) and *softly* (whole-word bias) is more likely to be present in the acoustic signal when the frequency of the derived form is relatively small compared to the frequency of the stem. In other words, constituent-driven processing favors the presence of [t], whereas word that are more independent of their constituents are more likely to be simplified phonologically (see also our discussion above of words such as *government*). Facts such as these show that there must be a trade-off of storage and computation in speech production as well.

Given the present results, it is interesting to revisit the status of the key exceptions to the Complexity Ordering hierarchy. In section 3 we raised the problem how to deal with suffix pairs that occur in both orders, as witnessed by pairs such as *-nessless* (*weaknessless*) and *-lessness* (*aimlessness*). Interestingly, the exceptional order *-nessless* is attested not only for sequences of two suffixes, but also for sequences of three suffixes, as shown by the following examples.¹¹

Thursday, May 1, 2008. **Eventfulnessless**. Guys, I was serious about my running out of ideas. If you wish to see a post up everyday, as I know many of you do and I thank you for your continued interest in my writing, SUBMIT IDEAS SO I HAVE STUFF TO WRITE ABOUT! <http://littlestsun.blogspot.com/2008/05/eventfulnessless.html>

JSTOR: A Study of Svatantrika ... tantric influence and repeatedly introduces tantric concepts such as primordial mind (sems-gnas-ma), **mindfulnessless** (dran-pa med-pa), and so forth. ... <http://ccbs.ntu.edu.tw/FULLTEXT/JR-PHIL/dreyfus1.htm>

¹¹We also found examples of adjectival *-ly* preceding the combination *-ness-less*:

The eternal knock against Qitely has been his complete **timelinessless**; monthly books become biennial with this man. DC's addressed this issue by making ASS ... <http://mesmerizationeclipse.blogspot.com/2007/01/this-month-in-supermen.html>

I can bring home the bacon ... — I dunno, **smellinessless** might be a good power if the people you're up against have acute sense of smell and therefore your lack of stink would mean they ... <http://potatohead.livejournal.com/327721.html>

Interestingly, these all involved cases in which the final suffix *-less* creates words that are used as nouns, instead of adjectives. Although this peculiarity may well have to do with processing difficulties (whereby the suffix may have lost its transpositional power), we do not have a real explanation to offer for this phenomenon.

The base words for these *-less*-formations have zero frequency in CELEX. This goes against the hypothesis of Complexity-Based Ordering, since under this hypothesis the base word would have to be more dependent on memory and in that sense less easily parsable. These formations constitute genuine counterexamples to approaches to Complexity-Based Ordering that do not allow leakage.

Similar conclusions can be reached by looking at the exceptional formations with the suffix pair *-lyAJ-ish* (cf. Table 4).

Join us at our **monthlyish** visit to the cinema! We're off to see Enchanted TWO weeks before it's official release! It's this Sunday, 2nd Dec ... <http://eventful.com/events/enchanted-at-the-barnet-odeon-/E0-001-007359946-4>

I'm A Poet, Do You Know It? Upright and righteous, **godlyish** people, We worshipped God in a place with a steeple. Humble and gentle, our motto was honesty, ... <http://myhome.spu.edu/lydia/poetry/>

I have a request ... — Zelda Universe Forums He is wearing a **knightlyish** brown thick tunic with patches here and there. His expression is slightly happy, but serious. He is armed with a sword with a ... <http://www.zeldauniverse.net/forums/artwork/49313-i-have-request.html>

The phonotactics of the two suffixes predicts that the sequence with the consonant-initial suffix, i.e. BASE-*ly*, will generally be more parsable than the sequence with the phonologically more integrated vowel-initial suffix *-ish* (cf. Hay 2003). The sequence *-lyAJ-ish* is therefore not only non-optimal in terms of Complexity-Based Ordering, but also in terms of phonotactics. Nevertheless, the non-optimal order is admitted by the selectional restrictions and it is in use.

A final and as yet unmentioned kind of exception involves repetition of the same suffix, which introduces the smallest possible cycle into the suffixal directed graph. In general, we disregarded combinations of the same suffix in the above because it has been claimed (e.g. in Hay & Plag (2004), footnote 3) that such formations are normally semantically ill-formed. There is, however, at least one suffix for which this claim does not hold, *-ish*. The following examples illustrate this.

What's up with Stephan Bonnar? I heard some random podcast where some guy (sorry I can't remember who but had a **jewishish** east coast accent) said something about bonnar treating h is ... <http://ninjashoes.net/forum/showthread.php?p=233454>

Paradise Place It's played with mischief, with fun but this is good fun, Theodosii has a **boyishishness** about him, maybe he's reliving an old memory, but now not as the boy ... http://www.paradiseplaceproductions.com/webpages/spassov_event.htm

For *-ish*, the reduplication is fully in line with its selectional restrictions, and the second example shows in addition how the reduplicated suffix serves as input to further suffixation with *-ness*.

What all these examples show is that although the graph of English suffix sequences tends to a surprising degree to acyclicity, there are various well-formed exceptions. This raises the question of how to understand the nature of these cycles from a theoretical point of view. Do the cycles arise from stochastic fluctuations in the rankings of suffixes on the hierarchy so that, for example, *-ness* is sometimes ranked higher and sometimes lower than *-less*, allowing both suffix orders to be realized? That is unlikely because then we would expect there to be many more cycles in our graphs since any two affixes should undergo similar random reorderings, contrary to fact.

As an alternative, one could think that the cycles emerge as the inevitable consequence of a fundamental property of human language, namely to have recursive structures. Although potentially interesting from a theoretical point of view, this does not seem to be a convincing explanation. We first have to distinguish between cyclicity and recursion. While an upward-pointing arrow is a sign of cyclicity, recursion does require a full loop back to the first suffix in a single formation. In English morphology it seems that the potential for recursion is there, but it is hardly ever realized. Examples from the web show that truly recursive formations do exist but seem to be carefully crafted and are increasingly hard to process.¹² Consider the following quotation from an e-mail signature (<http://mathforum.org/kb/message.jspa?messageID=252715&tstart=0>):

We must be fearless
We must have fearlessness
We must not be fearlessnessless
We must not have fearlessnesslessness
We must be fearlessnesslessnessless

This example shows that, like in syntactic processing, multiple recursion quickly leads to prohibitive comprehension difficulties. We therefore judge it to be unlikely that the cycles in our graphs exist specifically to subserve recursion. Instead, we submit that the cycles exist simply because the selectional restrictions of the suffixes involved fall out in this way.

Finally, it is noteworthy that the evidence for the directed graph for English suffixes being nearly acyclic is much stronger than the evidence for a correlation of Mean CO-Rank with suffix productivity, which, although significant, is not extremely strong ($R^2 = 0.33$). Likewise, the evidence for processing consequences of the Mean CO-Rank is much stronger than the evidence for its correlation with suffix productivity. This raises the question of what the results might be when even larger numbers of suffixes are taken into consideration. We expect the evidence for acyclicity to remain strong. However, it is conceivable that further research will show that for larger numbers of suffixes the correlation between productivity

¹²This can also be seen with prefixes like *great-*, as in *great-great-great-great-grandfather*, where one quickly loses track of the number of generations, or formations such as *anti-anti-missile-missile-missile*. (Thanks to Brian Joseph for these examples).

and mean CO-rank breaks down — recall that in the study of Hay & Plag (2004) with only 15 suffixes there was more support for correlational structure involving CO-rank and measures of constituent-driven processing. This possibility requires further reflection on the status of acyclicity as an independent processing principle in the grammar.

What other processing mechanism might drive the acyclicity that constrains the combinatorial possibilities of suffixes? An important aspect of human cognition in general, and of language processing in particular, is the ability to anticipate or plan for upcoming constituents (e.g. Hawkins & Blakeslee (2004)). When the listener has heard the first suffix of *restlessness*, and has deduced from the fine phonetic detail in the speech signal that another suffix is following (Kemps et al. 2005a,b), conditional probabilities may come into play. In the case of *restlessness*, the conditional probability of *-ness* given *-less*, or more generally,

$$(4) \quad \Pr(\text{suffix}_2|\text{suffix}_1),$$

describes the likelihood that the listener will correctly anticipate the second suffix, given the first one. In speech production, 4 captures the likelihood that the speaker will select the target suffix from the set of possible suffixes that might follow the first suffix.

We now note that this conditional probability can be much higher when the graph of suffix combinations is acyclic. In an acyclic graph, the number of suffixes that might potentially follow a given suffix decreases as one moves from lower-ranked to higher-ranked suffixes. In the ordered adjacency matrix shown in Figure 4, this is easily seen: As we move from *-th* down to *-wise* the number of affixes above the diagonal decreases. In other words, given that we know that the rank of the first suffix of the combination is, for example, 10, we know that the largest possible set of suffixes that might follow is constrained to the 21 suffixes to its right in the hierarchy. When the rank is 20, we know the continuation set contains only 11 candidates. The actual sets of possible continuation suffixes are, of course, further constrained by selectional restrictions, and contain only those suffixes that have a non-zero entry on the relevant rows in the ordered adjacency matrix (cf. Table 4). To obtain more precise estimates of the conditional probabilities as given in 4, we should not only take the number of attested affix combinations into account, but also weight these combinations for the number of different words instantiating these combinations, as well as the token frequencies of these words. Crucially, graphs that have very many cycles hardly constrain the set of possible continuation suffixes. In such graphs, it is impossible to restrict the set of suffixes that might follow to those that appear to the right in the hierarchy. In short, acyclic graphs afford enhanced anticipation (in comprehension) and enhanced planning (in production) compared to graphs with many cycles.

Interestingly, acyclicity may also link up with suffix productivity in a new way. Suffixes that occur to the left in the hierarchy (e.g., *-th*) are suffixes that provide the least effective conditional probabilities. For such suffixes, there are many possible continuation suffixes, and hence the probabilities of the individual suffixes all remain small. For suffixes to the right in the hierarchy (e.g., *-ness*), there are only few continuation suffixes, and as a consequence the conditional probabilities of these suffixes will tend to be much larger. In other words, the suffixes to the left in the hierarchy, the less productive suffixes, have a processing disadvantage compared to the suffixes to the right of the hierarchy, the more productive

suffixes. This change in processing complexity as we move from the left to the right of the hierarchy is, by logical necessity, fully correlated with the rank in the hierarchy.

If this understanding of why acyclicity is favored is correct, we also have an explanation of why the correlation with the \mathcal{P} measure of productivity is not particularly strong ($R^2 = 0.33$). The rank in the hierarchy captures (via conditional probabilities) the likelihood of the transition of a given suffix to a second suffix, whereas the \mathcal{P} measure estimates the overall likelihood of observing a new formation ending in the first suffix. The two likelihoods are correlated, but not identical.

A different question that we have not yet touched upon is to what extent the present findings might generalize to other languages with more productive morphology, such as Russian or Italian. We expect that in such languages the balance of storage and computation would shift dramatically in favor of parsing, and that parsing constraints would play a minor role at best. Given that in such languages the token frequencies of individual word forms are much reduced compared to English (as a much larger set of forms compete for the roughly the same usage space), the role of memory must be substantially reduced. Under these circumstances, CO-rank might well emerge as an independent principle uncorrelated with measures of constituent-driven processing.

Returning to English, a recurring comment that we have encountered is whether the very low-frequency forms that we have discussed are at all comprehensible. We have been told by native speakers (and professional linguists) that words such as *weaknessless* or *boyishishness* are uninterpretable. However, these words are used by other native speakers, and they are potential words of English in the sense that they do not violate any selectional restrictions. Why then do some native speakers have problems with these forms? We believe these problems arise from two sources. First, they involve the repetition of identical phonological/phonetic material across morphological boundaries, which is generally avoided in English (Stemberger 1981, Plag 1998, Raffelsiefen 1999). This avoidance phenomenon is known as morphological haplogy. Haplogical restrictions are common in English derivational morphology and may also have an influence here. Second, unease with these formations may arise due to processing difficulties. Speakers of English are used to encountering complex words that they have seen before. For *weaknessless* or *boyishishness*, which they have not seen before, speakers of English have to fall back on parsing which, as we have seen, plays a subordinate role in this language (as against languages with a rich morphology), and which is all the more difficult the more suffixes are to be processed.

Another issue that deserves mentioning is the effect of acoustic reduction on constituent-driven processing. Forms that undergo reduction may become less parsable during speech comprehension, especially when the suffix itself is substantially reduced. Interestingly, reduction affects higher-frequency words (Bell et al. 2003, Jurafsky et al. 2001) and affixes in higher-frequency words (Pluymaekers et al. 2005) most strongly, and may therefore be expected to be most pronounced for words with less productive affixes (Keune et al. 2005). In other words, reduction processes may foster entrenchment in memory notably at the lower end of the productivity spectrum, and at the same time render parsing processes more difficult. As a consequence, the present results, which are based on the written record and

on processing measures involving reading skills, may underestimate the role of memory in auditory lexical processing and speech production.

To summarize, our paper offers three key findings. First, suffixes can be ordered along a hierarchy according to which suffixes that are closer to the stem tend very strongly not to occur outside suffixes that occur not so close to the stem. This hierarchy was shown to correlate with the productivity of the suffixes. The findings of Hay and Plag (2004) stand the test of doubling the number of suffixes.

Second, behavioral data from lexical decision and word naming have shown that constituent-driven processing is not necessarily the most time-efficient way of processing. In particular, constituent-driven processing (as gauged in terms of rank in the hierarchy) does not stand in a linear relationship with processing costs. Rather, suffixes with *extreme* ranks in the Complexity Ordering hierarchy have lower processing costs than suffixes of medium rank. This can be interpreted as an effect of the opposing forces of storage and computation, with low-ranked suffixes profiting from whole-word storage and high-ranked suffixes from segmentation. This means that the model of Complexity-Based Ordering has been one-sided in its exclusive focus on the importance of constituent-driven processing and that it requires supplementation by a second and equally important focus on the role of memory.

Third, we have shown that the hierarchy of complexity-based ordering implies that the graph of suffix combinations is acyclic, and that the empirically observable near-acyclicity is very unlikely to be due to chance. Our hypothesis is that this acyclicity is functional for lexical processing as it allows for better estimates of transitional likelihoods from the first to the second suffix. These transitional likelihoods also link up with increasing productivity along the hierarchy. Future research has to investigate whether this new aspect of morphological productivity can be substantiated by behavioral and electro-physiological measures of lexical processing.

References

- Adams, Valerie. 2001. *Complex Words in English*. Harlow: Longman.
- Allen, Margaret R. 1978. *Morphological investigations*. Ph.D. thesis, University of Connecticut.
- Andrews, Sally A. 1992. Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, & Cognition* 18:234–254.
- Baayen, Harald. 2005. Data mining at the intersection of psychology and linguistics. Anne Cutler (ed.) *Twenty-first century psycholinguistics: Four cornerstones*, 69–83. Hillsdale, New Jersey: Erlbaum.
- Baayen, Harald. 2007. Storage and computation in the mental lexicon. Gonia Jarema & Gary Libben (eds.) *The Mental Lexicon: Core Perspectives*, (in press). Elsevier.

- Baayen, Harald. 2008. *Analyzing linguistic data. A practical introduction to statistics*. Cambridge: Cambridge University Press.
- Baayen, Harald & Anneke H. Neijt. 1997. Productivity in context: a case study of a Dutch suffix. *Linguistics* 35:565–587.
- Baayen, Harald & Fermín Moscoso del Prado Martín. 2005. Semantic density and past-tense formation in three Germanic languages. *Language* 81:666–698.
- Baayen, Harald & Robert Schreuder. 1999. War and peace: morphemes and full forms in a non-interactive activation parallel dual route model. *Brain and Language* 68:27–32.
- Baayen, Harald, Richard Piepenbrock & Leon Gulikers. 1995. *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.
- Baayen, Harald, Teun Dijkstra & Robert Schreuder. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 36:94–117.
- Baayen, Harald, Laurie B. Feldman & Robert Schreuder. 2006. Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* 53:496–512.
- Baayen, R. H., L. H. Wurm & J. Aycok. 2008a. Lexical dynamics for low-frequency complex words. a regression study across tasks and modalities. *The Mental Lexicon* 2:419–463.
- Baayen, R. Harald & Antoinette Renouf. 1996. Chronicling The Times: Productive Lexical Innovations in an English Newspaper. *Language* 72:69–96.
- Baayen, R. Harald, J. Davidson, Douglas & Douglas M. Bates. 2008b. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* Available online 3 March 2008.
- Balling, Laura & Harald Baayen. 2007. A study of lexical processing in Danish. *manuscript in preparation* .
- Balota, David A., Michael J. Cortese, Susan D. Sergent-Marshall, Daniel H. Spieler & Melvin J. Yap. 2004. Visual word recognition for single-syllable words. *Journal of Experimental Psychology:General* 133:283–316.
- Balota, David A., Melvin J. Yap, Michael J. Cortese, Keith I. Hutchison, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson & Rebecca Treiman. 2007. The english lexicon project. *submitted to Behavior Research Methods* 39(3):445–459. URL http://artsci.wustl.edu/~rtreiman/Selected_Papers/English_Lexicon_Project_userguide_in%20press.pdf.
- Barker, C. 1998. Episodic *-ee* in English: A Thematic Role Constraint on New Word Formation. *Language* 74(4):695–727.

- Bates, Douglas & Deepayan Sarkar. 2005a. *lme4: Linear mixed-effects models using Eigen and Eigenfaces*. R package version 0.95-8.
- Bates, Douglas M. 2005. Fitting linear mixed models in R. *R News* 5:27–30.
- Bates, Douglas M. & Deepayan Sarkar. 2005b. The lme4 library. [On-line], Available: <http://lib.stat.cmu.edu/R/CRAN/>.
- Bauer, Laurie. 1983. *English Word Formation*. Cambridge: Cambridge University Press.
- Bell, Allan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113:1001–1024.
- Bertram, Raymond, Harald Baayen & Robert Schreuder. 2000a. Effects of family size for complex words. *Journal of Memory and Language* 42:390–405.
- Bertram, Raymond, Robert Schreuder & Harald Baayen. 2000b. The balance of storage and computation in morphological processing: the role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Memory, Learning, and Cognition* 26:419–511.
- Booij, Geert E. 2002. Prosodic restrictions on affixation in dutch. Geert E. Booij & Jaap van Marle (eds.) *Yearbook of Morphology 2001*. Dordrecht: Kluwer.
- Brysbaert, Marc, Marielle Lange & Ilse Van Wijnendaele. 2000. The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology* 12:65–85.
- Burani, Cristina & Alfonso Caramazza. 1987. Representation and processing of derived words. *Language and Cognitive Processes* 2:217–227.
- Burnard, Lou. 1995. *Users guide for the British National Corpus*. Oxford university computing service: British National Corpus consortium.
- Burzio, Luigi. 1994. *Principles of English stress*. Cambridge: Cambridge University Press.
- Burzio, Luigi. 2002. Missing players: Phonology and the past-tense debate. *Lingua* 112:157–199.
- Bybee, J. L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.
- Chatterjee, Samprit, Ali S. Hadi & Bertram Price. 2000. *Regression analysis by example*. New York: John Wiley & Sons.

- Coltheart, Max, Eddy Davelaar, John T. Jonasson & Derek Besner. 1977. Access to the internal lexicon. S. Dornick (ed.) *Attention and performance*, volume VI, 535–556. Hillsdale, New Jersey: Erlbaum.
- Dalton-Puffer, Christiane & Ingo Plag. 2000. Categorywise, some compound-type morphemes seem to be rather suffix-like: On the status of -ful-, -type, and -wise in present day English. *Folia Linguistica* 34(3-4):225–244.
- De Vaan, L., Robert Schreuder & Harald Baayen. 2007. Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon* 2:1–23.
- Fabb, Nigel. 1988. English suffixation is constrained only by selectional restrictions. *Natural language and linguistic theory* 6:527–539.
- Faraway, Julian J. 2006. *Extending Linear Models with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Boca Raton, FL: Chapman & Hall/CRC.
- Fellbaum, Christiane (ed.) . 1998. *WordNet: An electronic database*. Cambridge, MA: The MIT Press.
- Gansner, Emden R., Eleftherios Koutsofios, Stephen C. North & Kiem-Phon Vo. 1993. A technique for drawing directed graphs. *IEEE Transactions on Software Engineering* 19(3):214–230.
- Giegerich, Heinz J. 1999. *Lexical strata in English: morphological causes, phonological effects*. Cambridge: Cambridge University Press.
- Harrell, Frank E. 2001. *Regression Modeling Strategies*. New York: Springer.
- Hawkins, Jeff & Sarah Blakeslee. 2004. *On intelligence*. New York: Henry Holt and Company.
- Hawkins, John. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hay, Jennifer B. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39:1041–1070.
- Hay, Jennifer B. 2002. From speech perception to morphology: Affix-ordering revisited. *Language* 78:527–555.
- Hay, Jennifer B. 2003. *Causes and Consequences of Word Structure*. New York and London: Routledge.
- Hay, Jennifer B. & Harald Baayen. 2002. Parsing and productivity. Geert E. Booij & Jaap Van Marle (eds.) *Yearbook of Morphology 2001*, 203–235. Dordrecht: Kluwer Academic Publishers.

- Hay, Jennifer B. & Ingo Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural language and linguistic theory* 22:565–596.
- Hayes, Bruce. 2001. Query Google. <http://www.linguistics.ucla.edu/people/hayes/QueryGoogle/>.
- Jespersen, Otto. 1942. *A Modern English Grammar. On Historical Principles. Part VI Morphology*. London: Allen and Unwin.
- Jungnickel, Dieter. 2007. *Graphs, networks and algorithms*. Berlin: Springer.
- Jurafsky, Daniel, Allan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. Joan L. Bybee & P. Hopper (eds.) *Frequency and the emergence of linguistic structure*, 229–254. Amsterdam: Benjamins.
- Kastovsky, Dieter. 1986. Productivity in word formation. *Linguistics* 24:585–600.
- Kemps, Rachel, Mirjam Ernestus, Robert Schreuder & Harald Baayen. 2005a. Prosodic cues for morphological complexity: The case of Dutch noun plurals. *Memory and Cognition* 33:430–446.
- Kemps, Rachel, Lee H. Wurm, Mirjam Ernestus, Robert Schreuder & Harald Baayen. 2005b. Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes* 20:43–73.
- Keune, Karen, Mirjam Ernestus, R. Van Hout & Harald Baayen. 2005. Social, geographical, and register variation in Dutch: From written ‘mogelijk’ to spoken ‘mok’. *Corpus Linguistics and Linguistic Theory* 1:183–223.
- Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. H. Van der Hulst & N. Smith (eds.) *The structure of phonological representations*, 131–176. Dordrecht: Foris.
- Kostić, Aleksandar, Tanja Marković & Aleksandar Baucal. 2003. Inflectional morphology and word meaning: Orthogonal or co-implicative domains? Harald Baayen & Robert Schreuder (eds.) *Morphological Structure in Language Processing*, 1–44. Berlin: Mouton de Gruyter.
- Krott, Andrea, Robert Schreuder & Harald Baayen. 1999. Complex words in complex words. *Linguistics* 37(5):905–926.
- Landauer, Thomas K. 1986. How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science* 10:477–493.
- Levelt, William J. M., A. Roelofs & A. S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22:1–38.

- Lewis, Michael B., S. Gerhand & H. D. Ellis. 2001. Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects? *Cognition* 78:189–205.
- Ljung, Magnus. 1970. *English Denominal Adjectives: A Generative Study of the Semantics of a Group of High-frequency Denominal Adjectives in English*. Stockholm: Almqvist & Wiksell.
- Luce, Paul A. 1985. *Structural distinctions between high and low frequency words in auditory word recognition*. PhD. dissertation, Indiana University.
- Malkiel, Yakuv. 1977. Why *ap-ish* but *worm-y*? Paul J. Hopper (ed.) *Studies in Descriptive and Historical Linguistics*, 341–364. Amsterdam: Benjamins.
- Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word Formation. A Synchronic-Diachronic Approach*. München: Beck'sche Verlagsbuchhandlung.
- Miller, George A. 1990. Wordnet: An on-line lexical database. *International Journal of Lexicography* 3:235–312.
- Mohanan, Karuvannur P. 1986. *The theory of lexical phonology*. Dordrecht: Reidel.
- Morrison, Catriona M. & Andrew W. Ellis. 2000. Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology* 91:167–180.
- New, Boris, Ludovic Ferrand, Christophe Pallier & Marc Brysbaert. 2006. Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review* 13(1):45–52.
- OED. 1994. *The Oxford English Dictionary (2nd edition), on Compact Disk*. Oxford: Oxford University Press.
- Pinheiro, Jose C. & M. Bates, Douglas. 2000. *Mixed-effects models in S and S-PLUS*. Statistics and Computing. New York: Springer.
- Pinker, Steven. 1991. Rules of language. *Science* 153:530–535.
- Pinker, Steven. 1997. Words and rules in the human brain. *Nature* 387:547–548.
- Pinker, Steven. 1999. *Words and Rules: The Ingredients of Language*. London: Weidenfeld and Nicolson.
- Pisoni, David B., Howard C. Nusbaum, Paul A. Luce & Lousia M. Slowiaczek. 1985. Speech perception, word recognition and the structure of the lexicon. *Speech Communication* 4:75–95.
- Plag, Ingo. 1996. Selectional restrictions in English suffixation revisited: a reply to Fabb (1988). *Linguistics* 34:769–798.

- Plag, Ingo. 1998. Morphological haplology in a constraint-based morpho-phonology. Wolfgang Kehrein & Richard Wiese (eds.) *Phonology and Morphology of the Germanic Languages*, 199–215. Tübingen: Niemeyer.
- Plag, Ingo. 1999. *Morphological productivity: structural constraints in English*. Berlin: Mouton de Gruyter.
- Plag, Ingo. 2002. The role of selectional restrictions, phonotactics and parsing in constraining suffix ordering in English. *Yearbook of Morphology 2001* 285–314.
- Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
- Pluymaekers, Mark, Mirjam Ernestus & Harald Baayen. 2005. Frequency and acoustic length: the case of derivational affixes in Dutch. *Journal of the Acoustical Society of America* 118:2561–2569.
- Moscoso del Prado Martín, Fermín, Raymond Bertram, Tuomo Haikio, Robert Schreuder & Harald Baayen. 2004a. Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition* 30:1271–1278.
- Moscoso del Prado Martín, Fermín, Aleksandar Kostić & Harald Baayen. 2004b. Putting the bits together: An information theoretical perspective on morphological processing. *Cognition* 94:1–18.
- R Development Core Team. 2005. *R Language Definition. Version 2.4.0*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- R Development Core Team. 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Raffelsiefen, Renate. 1999. Phonological constraints on English word-formation. Geert E. Booij & Jaap van Marle (eds.) *Yearbook of Morphology 1998*, 225–287. Dordrecht: Kluwer.
- Riddle, Elizabeth. 1985. A historical perspective on the productivity of the suffixes *-ness* and *-ity*. Jacek Fisiak (ed.) *Historical Semantics, Historical Word-Formation*, 435–461. New York: Mouton.
- Ryder, Mary E. 1999. Bankers and blue-chippers: an account of -er formations in Present-day English. *English Language and Linguistics* 3(2):269–297.
- Saciuk, Bogdan. 1969. The stratal division of the lexicon. *Papers in Linguistics* 1:464–532.
- Scarborough, Rebecca A. 2004. Coarticulation and the structure of the lexicon. Ph.D. thesis, University of California. UCLA dissertation.

- Schreuder, Robert & Harald Baayen. 1997. How complex simplex words can be. *Journal of Memory and Language* 37:118–139.
- Schreuder, Robert, Cristina Burani & Harald Baayen. 2003. Parsing and semantic opacity. Egbert M. H. Assink & Dominiek Sandra (eds.) *Reading Complex Words. Cross-language studies*, 159–189. Dordrecht: Kluwer.
- Seidenberg, Mark S. 1987. Sublexical structures in visual word recognition: Access units or orthographic redundancy. Max Coltheart (ed.) *Attention and Performance XII*, 245–264. Hove: Lawrence Erlbaum Associates.
- Selkirk, Elisabeth. 1982. *The Syntax of Words*. Cambridge: The MIT Press.
- Siegel, Doris. 1974. Topics in English morphology. Ph.D. thesis, MIT. Ph. D. thesis, MIT.
- Spencer, Andrew. 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Cambridge: Cambridge University Press.
- Taft, Marcus. 1979. Recognition of affixed words and the word frequency effect. *Memory and Cognition* 7:263–272.
- Taft, Marcus. 1988. A morphological decomposition model of lexical representation. *Linguistics* 26:657–667.
- Taft, Marcus. 2004. Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology* 57A:745–765.
- Taft, Marcus & Kenneth I. Forster. 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior* 15:607–620.
- Van Heuven, Vincent J., Anneke Neijt & Maarten Hijzeldoorn. 1994. Automatische indeling van nederlandse worden op basis van etymologische filters. *Spektator* 23:279–291.
- Van Son, Rob J.J.H. & Louis C.W. Pols. 2003. Information Structure and Efficiency in Speech Production. *Proceedings of Eurospeech-2003*, 769–772. Geneva, Switzerland.
- Van Son, Rob J.J.H. & Jan P.H. Van Santen. 2005. Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication* 47:100–123.
- Vitevitch, Michael S. 2002. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 28(4):735–747.
- Vitevitch, Michael S. & Melissa K. Stamer. 2006. The curious case of competition in Spanish speech production. *Language and Cognitive Processes* 21(6):760–770.

- Wright, Richard. 2004. Factors of lexical competition in vowel articulation. John Local, Richard Ogden & Temple Rosalind (eds.) *Papers in Laboratory Phonology 6*, 75–87. Cambridge: Cambridge University Press.
- Wurm, Lee H. 1997. Auditory Processing of Prefixed English Words is Both Continuous and Decompositional. *Journal of Memory and Language* 37:438–461.
- Wurm, Lee H. 2000. Auditory Processing of Polymorphemic Pseudowords. *Journal of Memory and Language* 42:255–271.
- Wurm, Lee H. & Joanna Aycock. 2003. Recognition of spoken prefixed words: the role of early conditional root uniqueness points. Harald Baayen & Robert Schreuder (eds.) *Morphological Structure in Language Processing*, 259–286. Berlin: Mouton de Gruyter.
- Zipf, George K. 1949. *Human Behavior and the Principle of the Least Effort. An Introduction to Human Ecology*. New York: Hafner.

Appendix

Examples of attested English derived words with two-suffix combinations (from OED, if not indicated otherwise).

Suffixes studied by Hay & Plag (2004).

lengththen depthless flattener (internet) *flattenee* (internet) *preacherling breweress loverly printerdom loverhood controllership robberish leaderless tumblerful* (BNC) *saplinghood ducklingship seedlingless refugee-ess* (internet) *employeeshood assigneeship princessly* (internet) *princessdom priestesshood governessship governessless knightlyhood woollyish kingdomless kingdomful courtliness childhoodless* (internet) *ensorshipless* (internet) *kinshipful amateurishness aimlessness carefulness*

Suffixes added in this study.

meteorette protectorian factorage detectorist traitorous actorism scissorwise rhymsterette spinsterian hucksterage spinsterous hucksterism activist activism parliamentary (adjective) *parliamentary* (noun) *developmentist garmentry* (BNC) *medicamentous sacramentism pedimentwise librarian voluntaryist apothecariry contrarious secretaryism contrariwise briquetage novelettist suffragettism physicianary guardianage Europeanist sylvanry ruffianous Europeanism Christianwise manifoldwise militaryment artillerist voluntaryism plumagery umbrageous percentagewise dentistry effronterist mysterious adaptively adventurously complementarily*

Suffixes of Hay & Plag (2004) combined with suffixes added in the present study

depthwise lengthenment sleeperette farmerage comsumerist loverwise boxerism grovellingwise absenteeism huntress-wise beastlywise orderlyism foolishment Irishian Yiddishist Scottishry foolishwise Britishism careleswise despiteful-wise courtshipment businesswise (BNC) *neighbourhoodism* (BNC) *despitefully babyishly blushlessly*

Suffixes added in the present study combined with suffixes of Hay & Plag (2004)

tenfoldness professorling editoress sailorly protectordom traitorhood administratorship spectatorish connectorless traitorful spinsterly gansterdom spinsterhood hucksteress tapstership spinsterish addictiveness governmentship documenter experimentee secretaryhood justiciaryship dictionaryless noveletteish musicianer barbarianess guardianly christiandom ruffianhood musicianship christianish guardianless victorianness villagedom villagehood villageship carriageless foolageness carriageful elementariness alchemister artistess artistly artistdom evan-

gelistship nurserydom ministryship otherwiseness