The directed compound graph of English
An exploration of lexical connectivity and its processing consequences

R. Harald Baayen
University of Alberta
e-mail: baayen@ualberta.ca

March 29, 2010

Abstract

This study explores the consequences of morphological connectivity for English compounds, combining tools from graph theory with measures of lexical processing costs as available in the English Lexicon Project (Balota et al., 2007). The directed compound graph reveals a significant trend to acyclicity just as the directed affix graphs of Hay and Plag (2004); Plag and Baayen (2009); Zirkel (2010), and similar correlations of rank and productivity. Rank in the directed graph, however, fails to correlate with measures of processing complexity. In order to understand the high degree of acyclicity, it is hypothesized that the activation of more distant neighbors in the lexical network is disadvantageous. A measure for more distant lexical neighbors, secondary family size, is proposed, and shown to have an inhibitory effect in visual lexical decision and word naming. Furthermore, an inhibitory effect of the shortest path from head to modifier is documented, and shown to depend on a specific time window within which activation reaching the modifier disrupts the process of compound interpretation.

keywords directed graph, strongly connected component, productivity, lexical decision, naming, mediated priming, complexity-based ordering

1 Introduction

Recent studies of derivational morphology have documented that sequences of suffixes (Hay and Plag, 2004; Plag and Baayen, 2009) and prefixes Zirkel (2010) can, with remarkably few exceptions, be ordered in an acyclic directed graph. That is, given a set of suffixes $\{S_1, S_2, \ldots, S_n\}$, an ordering can be found such that for any complex word ending in the derivational suffix S_i followed by the derivational suffix S_j, S_i precedes S_j in the ordering for any i and j. This ordering was argued by Hay and Plag (2004) to arise due to processing complexity, with less productive and less parsable affixes occurring closer to the base, and has been proposed as a *complexity-based* ordering. For suffixes, the partial ordering also follows largely from selectional restrictions (Hay and Plag, 2004), but for prefixes the selectional restrictions allow for many more prefix combinations than actually attested (Zirkel, 2010). In their study of 31 English suffixes, Plag and Baayen (2009) failed to obtain strong evidence supporting the original hypothesis that the acyclicity of the directed suffix graph is driven by the relative difficulty of parsing constituent sequences.

For compounds, selectional restrictions and junctural phonotactics do not play the important roles documented for English derivation. If we can establish that nevertheless the modifier-head pairs of compounds show the same degree of acyclicity that characterizes derivation, then this will provide further evidence that the trend towards acyclicity is not due to solely, or primarily, processing constraints on the first and second element in sequences of constituents.

If the ordering of constituents in compounds is not complexity-based, the surprising degree of acyclicity charaterizing constituent sequences in morphological networks remains to be explaned. This study explores an explanation based on the hypothesis that the co-activation of more distant morphological relatives due to spreading activation is disadvantageous for lexical processing.

In what follows, we first examine the directed compound graph, the extent to which it is acyclic, and whether the ordering of constituents in the graph can be linked to processing complexity. We then consider in more detail the connectivity in the graph, with as tools the graph-theoretical concept of the strongly connected component, and a new measure for the amount of more distant lexical connectivity, the secondary family size. Finally, we zoom in on the strongly connected component of the compound graph to address the question of whether activation spreading from the head back to the modifier affects lexical processing.

2 The directed compound graph

Is the English compound directed graph more or less acyclic, with a relatively small proportion of exceptions comparable to the proportion of exceptions observed for English derivation? To answer this question, we extracted 3880 two-constituent compounds with monomorphemic nouns as base words from the CELEX lexical database (Baayen et al., 1995). These 3880 compounds jointly comprised 2200 different base words. The DOT representation (Gansner et al., 1993) of the corresponding directed graph (2200 nodes, 3880 edges), obtained with the `Rgraphviz` package in `R` (Gentry et al., 2009), revealed 325 exceptions to acyclicity. The rate of exceptions, $325/3880 = 0.084$, is comparable to the exception rate reported by Plag and Baayen (2009) for 31 English derivational suffixes, $10/161 = 0.062$ ($X^2_{(1)} = 0.69, p > 0.4$). This indicates that a clear trend towards acyclicity characterizes not only suffixal derivation and prefixal derivation (Zirkel, 2010) but also compounding in English.

In order to assess whether the observed trend towards acyclicity constitutes reason for surprise, we proceeded as follows. In a completely acyclic adjacency matrix, all nonzero entries can be ordered above the main diagonal, in which case the matrix is in upper diagonal form. If the nonzero entries in the upper triangle of the matrix are uniformly distributed across this upper triangle, the row sums and the column sums are negatively correlated. This is easily seen for an adjacency matrix for which all entries above the main diagonal are 1, and all entries on and below the diagonal are zero. For an n by n matrix, the row sums are $n-1$, $n-2$, $n-3$, ..., 2, 1, 0, while the column sums are 0, 1, 2, 3, ..., $n-2$, $n-1$, yielding a perfect negative correlation ($r = -1$). As the adjacency matrix becomes more sparse, this negative correlation will be masked by more noise. However, compared to random matrices with the same sparseness, the observed correlation should remain in the extreme of the distribution of correlations.

Importantly, for evaluating whether an empirical adjacency matrix approaches acyclicity, it is not necessary to bring the adjacency matrix in upper diagonal form, which is an NP-hard problem: The correlation of row sums and column sums remains unchanged when column i and j are exchanged simultaneously with exchanging rows i and j, the basic operation for bringing the adjacency matrix in upper diagonal form.

The observed Spearman rank correlation for the observed adjacency matrix is -0.134. The range of correlations obtained by independently permuting rows and columns of the adjacency matrix 1000 times (thereby completely randomizing affix orders) was [-0.063, 0.073], indicating that the probability of observing the actually observed, and more extreme, correlation by chance is less than 0.001.

The observed trend towards acyclicity raises the question of whether there is a relation between rank (vertical position in the graph) and constituent productivity, similar to the relation between affix productivity and rank reported for English derivation by Hay and Plag (2004), Plag and Baayen (2009) and Zirkel (2010). We approach this question by gauging separately the productivity of the first constituent (the modifier) and the productivity of the second constituent (the head).

The measure of affix productivity that emerges as most robustly correlated with rank from the abovementioned studies on affix productivity is the category-conditioned degree of productivity, the ratio of hapax legomena with a given suffix and the total number of tokens with that suffix. Unfortunately, the CELEX lexical database does not provide reliable information on hapax legomena. We therefore assessed constituent productivity through the type count of the compounds sharing the first constituent (the modifier family size) and similarly the count of compounds sharing the second constituent (the head family size). The choice for these family size counts is further motivated by the following considerations. First, the modifier and head constituent families of compounds are known to form the domains of probabilistic generalization for interfixes in Dutch and German (Krott et al., 2001, 2004, 2007), for the interpretation of novel and existing compounds (Gagné and Shoben, 1997; Gagné, 2001), and for the assignment of compound stress (Plag et al., 2007; Plag, 2010; Plag and Kunter, 2010; Arndt-Lappe, 2010). Second, it is known that in visual comprehension the modifier and head family sizes are co-determinants of the time spent by the eye on a given constituent (Kuperman et al., 2008, 2009).

Of the 2200 compound constituents, 710 are used exclusively as head, 902 are used exclusively as modifier, and 588 occur both as head and as modifier (e.g., *soup*, as in *pea soup* and *soup kitchen*). Words that are used both as modifier and as head, unsuprisingly, are somewhat more frequent than words that are used either as head or as modifier (mean log frequencies 6.78 and 5.12, $t(1437.375) = 20.74, p < 0.0001$). The words that appear both as head and as modifier are the source of the violations of acyclicity in the directed compound graph: The 325 exceptional compounds comprise 225 distinct base words, all of which occur both as head and as modifier. Note that functioning both as modifier and as head, although a necessary condition, is not a sufficient condition for giving rise to exceptional compounds. For instance, *soup* in the chain *pea soup* → *soup kitchen* does not contribute to the set of exceptional compounds.

For the evaluation of the relation between productivity and ranking in the nearly acyclic compound graph, consider that in order to minimize the number of exceptions to acyclicity, it is necessary to order those words that occur only as modifier higher up in the graph, and those words that occur only as head lower down in the graph. The more productive a word is as modifier, the more important it will be to order it higher in the graph. Similarly, the more productive a word is as a head, the more crucial it will be to order it lower in the graph. For the words that take on both functions, some compromise solution will have to be found. In what follows, we use the coordinate system of the graph obtained with the DOT layout of the Gansner et al. (1993) algorithm using `plot.graphNEL` in the `Rgraphviz` package. In this graph, which is too complex to reproduce here, words occurring high up in the graph have large Y-coordinates, whereas words occurring lower in the graph have Y-coordinates closer to zero. Following Hay and Plag (2004), larger Y-coordinates correspond to lower ranks, and smaller Y-coordinates to greater ranks. Stated in terms of Y-coordinates, we therefore expect to find that a greater modifier productivity correlates positively with the Y-coordinates, and that a greater head productivity correlates negatively with the Y-coordinates. These correlations with constituent productivity reverse when phrased in terms of ranks. (In what follows, each distinct Y-coordinate is assigned a distinct rank. Words with the same Y-coordinate therefore share the same rank.)

	Estimate	Std. Error	t value	p value
Head: intercept	50.5113	1.4183	35.6138	0.0000
Head: linear	31.7507	2.9998	10.5843	0.0000
Head: spline	-25.7197	4.0007	-6.4289	0.0000
Modifier: intercept	64.9062	1.3020	49.8516	0.0000
Modifier: linear	-4.7250	1.3767	-3.4321	0.0006

Table 1: Coefficients for two linear models, one for the head, and one for the modifier, fitting the Y-coordinate to log constituent family size.

The expected correlations are observed for both head and modifier, when the Y-coordinate of a head (or modifier) is regressed on the log of the positional family size. The positional family size of a compound's constituent is the number of compounds in which that constituent occurs in the same functional position (modifier c.q. head). In this study, positional family sizes were calculated on the basis of the compounds in the present sample, i.e., they are conditional on the modifier or head being a constituent of a bimorphemic noun-noun compound. For the head, a negative decelerating curve was observed, modeled with a restricted cubic spline (Harrell, 2001) with three knots, such that for the higher log head family sizes, no further decrease in Y coordinate was present. For the modifier, the Y-coordinate

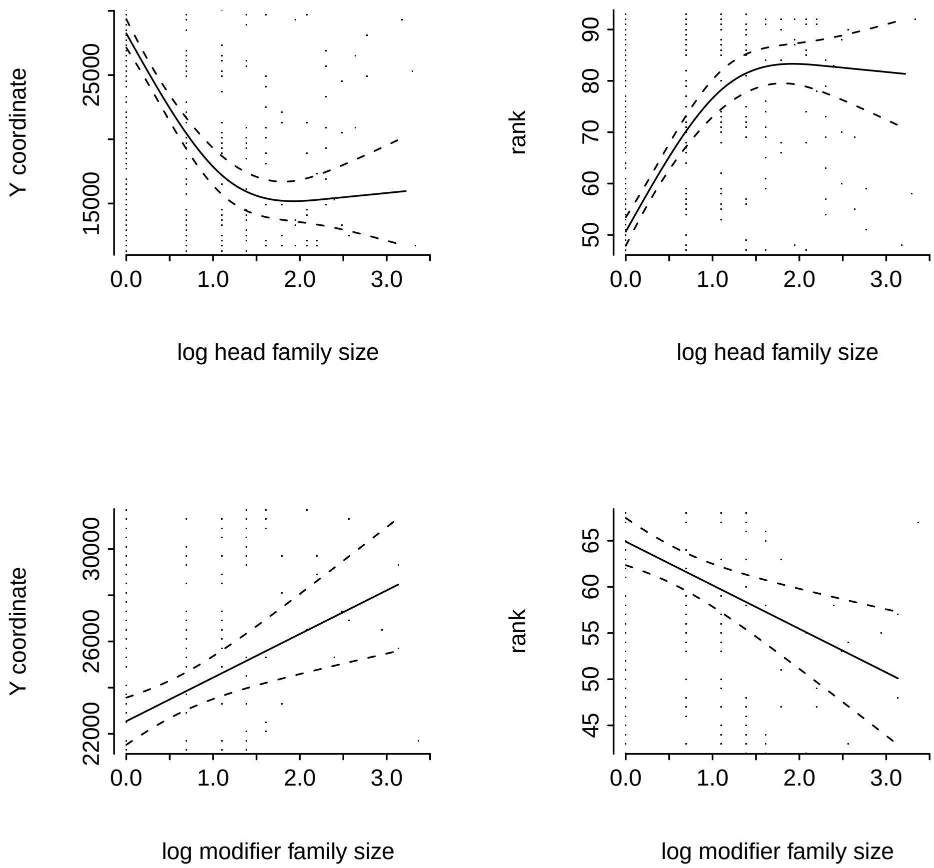

Figure 1: Predicted Y-coordinate (left panels) and rank (right panels) in the directed compound graph for the head (upper panels) and modifier (lower panels) constituents, obtained with four separate regression models.

increased linearly with log family size. Table 1 lists the coefficients of these models, and Figure 1 visualizes the models fitted to the Y-coordinates (left panels) and ranks (right panels) for heads (upper panels) and modifiers (lower panels).

The positive slope for rank that characterizes heads with a log family size less than 1.5 (1106 of the 1298 observations) replicates the positive slopes reported by Hay and Plag (2004) and Plag and Baayen (2009) for derivational suffixes. For the modifiers, the negative slope of rank replicates the study by Zirkel on prefixes in English (Zirkel, 2010). Zirkel observed a positive correlation of rank and (hapax-conditioned) degree of productivity for 15 English prefixes, with the lower rank assigned to the inner prefix, and the higher rank to the outer prefix. This approach mirrors the situation for suffixes, where the lower rank is also assigned to the inner suffix, and the higher rank to the outer suffix. In other words, ranks are traditionally assigned from the center to the periphery, to the left for prefixes, and to the right for suffixes.

When ranks are brought in line with linear order, from word beginning to word end, then the ranks for the prefixes have to be reversed. The correlation between this new "linear rank" for prefixes and prefix productivity is negative, just as the correlation between modifier productivity and rank is negative. Unlike for derivation, the distinction between bound and free forms does not come into play. When we focus on the modifier and its rank and productivity, the modifier precedes the head, and behaves exactly like a prefix in a prefix sequence with respect to the correlation of rank and productivity. When we focus on the head, which follows the modifier, it behaves exactly like a suffix in a sequence of suffixes. Across affixation and compounding, the same principles are at work. In order to approximate acyclicity, constituents that are productive and *precede* other constituents have to be ordered high in the graph, while constituents that are productive and *follow* other constituents have to be ordered low in the graph.

The existence of approximate acyclicity, combined with the significant correlations of ranks with degrees of constituent productivity, raises the question of whether it is at all surprising to find a correlation between productivity and rank. On the one hand, this correlation is not surprising, in the sense that if there are differences in productivity, and if it is possible to approach acyclicity, then this correlation is present by necessity. On the other hand, the mere fact that it is possible to approach acyclicity is genuinely surprising, and highly unlikely to arise under chance conditions.

When acyclicity was first observed, it was hypothesized that the ordering in the graph is motivated by processing constraints. Hay (2002) argued that more parsable affixes should be peripheral to less parsable affixes, where parsability was related to relative frequency (Hay, 2001), semantic transparency, and junctural phonotactics. Hay and Plag (2004) observed for 15 derivational suffixes that rank in the acyclic graph correlated not only with productivity (as measured by hapax-conditioned degree of productivity, type parsing ratio, token parsing ratio) but also with phonological boundary strength.

However, two follow-up studies provide evidence only for a correlation of rank with hapax-conditioned degree of productivity: a survey of 31 suffixes of Plag and Baayen (2009), as well as a study of 15 prefixes by Zirkel (2010). Parsing ratios and phonological measures of junctural phonotactics did not reach significance in these studies. For the compounds presently under study, there is also no correlation of rank with the frequency of the biphone straddling the boundary of modifier and head. Furthermore, almost all the compounds in our data set are highly parsable in the relative frequency sense of Hay (2001). For the subset of the data for which we have reliable frequency information (see below), compound frequency is greater than the frequency of the head in only 10 cases out of 1252, and there are no cases where the compound frequency exceeds both the frequency of its modifier and that of its head. As a consequence, the role of relative frequency is much reduced for compounds compared to derived words. In the dual route model of Hay (2001), this would indicate a strong parsing bias for allmost all compounds, and a ceiling effect for productivity and rank. Although indeed a ceiling effect characterizes the correlation of rank and head productivity, this ceiling effect is observed only for a minority of very productive heads. We therefore conclude that the presence of a statistically surprising degree of acyclicity for compounds indicates that the trend towards acyclicity and the correlation of rank and degree of productivity are motivated, at least for compounding, independently of parsability.

Acyclicity may offer processing advantages other than local parsability. Plag and Baayen (2009) speculate that acyclicity may be advantageous for predicting upcoming constituents: Given that the current constituent has rank R, all constituents with rank $r < R$ can be ruled out as possible upcoming constituents in the word. However, the present compound graph indicates, thanks to the large number of observations on which it is based, that there is a non-negligible number of compounds that contribute cycles to the graph. This raises the question of whether such highly connected compounds are more difficult to process than normal compounds. If so, this would provide straightforward support for the hypothesis that cycles are in some way computationally costly. The alternative, a processing advantage for highly connected compounds, would indicate that cycles in the compound graph have their own processing advantage to offer, leading to a system in which the high-level global advantages of acyclicity are balanced against the local, low-level advantages of being part of a cycle. In what follows, we explore these two alternatives by inspecting the strongly connected components of the compound graph.

3 The strongly connected component of the directed compound graph

A strongly connected component of a directed graph is a subgraph such that any node in that subgraph can be reached from any other node. By inspecting the number and size of the strongly connected components of the compound graph, enhanced insight can be obtained into how and why acyclicity is violated. It turns out, using the `strongComp` function (implementing Tarjan's algorithm) in the `RGBL` package (Carey et al., 2009), that the compound graph contains one large strongly connected subgraph, comprising 344 vertices, as well as one small strongly connected subgraph consisting solely of *price* and *list*, the constituents of the compounds *price list* and *list price*. The 344 vertices in the non-trivial strongly connected subgraph support 983 compounds, about one fourth of the total number of compounds. All but four of the exception constituents participate in the strongly connected component, *agar* (in *agar-agar*), *hula* (in *hula-hula*), and the abovementioned *price* and *list*. As we shall see below, it is being part of the strongly connected component rather than just being exceptional with respect to acyclicity that has consequences for lexical processing.

For studying the processing consequences of membership in the strongly connected component, we need to add information to our database concerning the lexical distributional properties of the compounds and their constituents, as well as measures of processing complexity. Lexical distributional information was extracted from the CELEX lexical database. As counts in CELEX are string-based, no information is available about spaced compounds. Consequently, the analyses to follow are all based on compounds written in CELEX as one word, or written with a hyphen. We further imposed the restriction that a compound should be listed with non-zero frequency. Processing information was extracted from the English Lexicon Project (Balota et al., 2007). In all, 1252 compounds were available both in CELEX and in the English Lexicon Project. Of these compounds, 830 were not exceptional and not part of the strongly connected component (e.g., *airbase*), 242 were exceptional and part of the strongly connected component (e.g., *armchair*), and 180 were both exceptional and part of the strongly connected component (e.g., *baseline*).

For the statistical analysis, a Generalized Additive Model, henceforth GAM, was used. GAMs provide a more flexible and precise way of modeling interactions involving two (or more) numerical predictors than the standard linear model. A generalized additive model consists of two parts, a parametric part identical to that of standard linear models, and a non-parametric part that provides non-parametric functions for modeling wiggly surfaces in two or higher dimensions. In the present study, we make use of so-called tensor products to model such surfaces, recommended by Wood (2006) for data with non-isotropic predictors. Tensor product functions are non-parametric in the sense that we will not be interested in the parameters that these smoothing functions use internally, but only in how well the shape of a given surface is captured. When fitting a tensor smoother to the data, it is important to avoid both undersmoothing and oversmoothing. We have used the default of the `gam` implementation of the MGCV package of Wood (2006) (version 1.4-1.1), which estimates the optimal smoothness from the data using generalized cross validation. The greater the estimated degrees of freedom (edf) for a tensor product term, the more the smoother invests in wiggliness. For other examples of GAMs applied to compound processing, see (Baayen et al., 2010).

A generalized additive model (Wood, 2006) fitted to the lexical decision latencies revealed linear facilitation for compound frequency, modifier frequency, and modifier family size. Heads that are also attested as modifiers in our database enjoyed a small processing advantage as well. These partial effects of the linear terms of the generalized additive model are visualized in Figure 2. (Due to the identifiability constraints in generalized additive models, confidence intervals fan out from zero.) The corresponding coefficients and associated statistics are listed in Table 2. We discuss the nonlinear part of the model below.

The (log-transformed) frequency of the modifier and log modifier family size are sig-

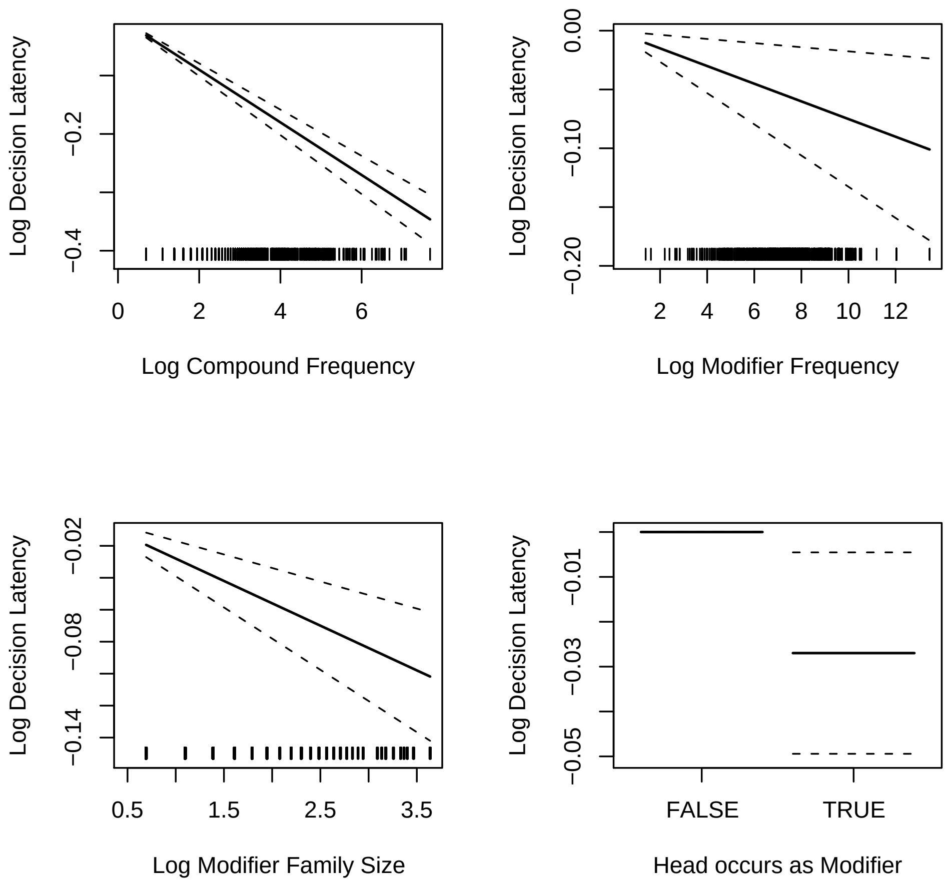

Figure 2: Lexical decision latencies in the English Lexicon Project as a function of log compound frequency, log modifier frequency, log modifier family size, and a binary factor for whether the head is also used as modifier, fitted with a generalized additive model.

nificantly correlated, $r = 0.59$, but the collinearity of the design matrix is small enough ($\kappa = 16.4$, Belsley et al. (1980)) and the data set large enough that no further corrective measures are required. Head frequency and head family size are also significantly correlated ($r = 0.65$), but here inclusion of either measure renders the other measure non-significant. In what follows, we work with the family size measure, as this is the measure that is most directly related to the structure of the compound directed graph, but it should be kept in mind that an equivalent model can be obtained by replacing the family size measure by the head frequency measure.

	Estimate	Std. Error	t value	p value
Intercept	6.9208	0.0203	341.6953	0.0000
Log Modifier Family Size	-0.0280	0.0055	-5.0643	0.0000
Compound Frequency	-0.0451	0.0027	-16.5240	0.0000
Modifier Frequency	-0.0075	0.0029	-2.6106	0.0091
Head is also used as Modifier	-0.0270	0.0112	-2.4039	0.0164

Table 2: Coefficients for the linear predictors in the generalized additive model fitted to the log-transformed lexical decision latencies.

Head family size emerged as significant in a three-way interaction (modeled with a tensor product) with whether the head is part of the strongly connected component of the compound graph, as well as in interaction with a new family size measure, the compound's (log-transformed) secondary family size. The secondary family size is obtained by summing, across both constituents, the positional family sizes of their compound family members. For instance, *trolley* occurs as modifier in *trolley car* and *trolleybus*, and as head in *tea-trolley*. The number of compounds in which *tea* occurs either as head or as modifier is 25, and the corresponding counts for *car* and *bus* are 16 and 3. Thus the total secondary family count is 44. As these counts include the head and modifier primary family sizes, the measure that we used in our model was the residual of the log-transformed secondary family size count regressed on the log-transforms of the primary family size counts for head and modifier. The resulting measure reflects the connectivity of a compound in the compound graph, in as far as that connectivity is not carried by the immediate connectivity of the modifier and head themselves.

Figure 3 presents contour plots for the fitted surface for the decision latencies predicted from log head family size, (residualized) log secondary family size, and membership in the strongly connected component, modeled with a tensor product. Darker shades of gray represent shorter latencies. The top panel represents the compounds with heads that are not part of the strongly connected component (modeled with a tensor product with 16.346 edf, $F(16.846, 1220.989) = 2.946, p < 0.0001$), the bottom panel shows the corresponding surface for the compounds in the strongly connected component (modeled with a tensor product with 8.665 edf ($F(9.165, 1200.989) = 4.864, p < 0.0001$). Likelihood ratio tests comparing this model with simpler models supported the model with the three-way interaction.

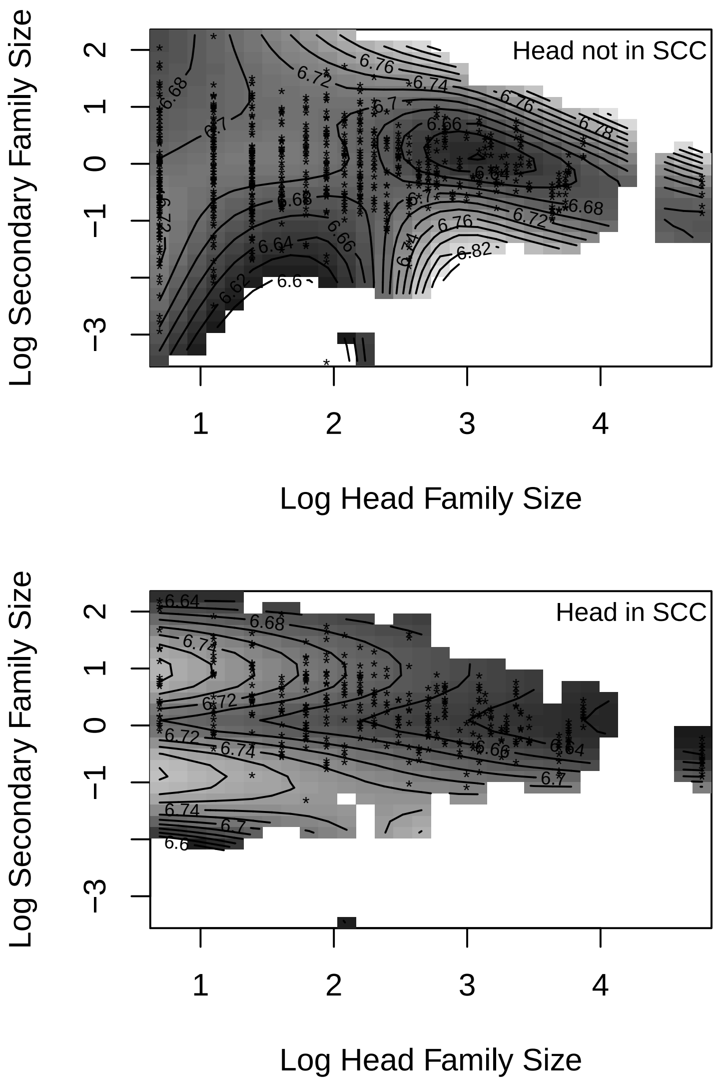

Figure 3: Fitted lexical decision latencies in the English Lexicon Project as a function of log head family size and Residualized log secondary family size, for compounds not in (upper panel) and in (lower panel) the strongly connected component, fitted by a generalized additive model. Contour lines connect identical predicted log latencies. Lighter shades of gray represent longer latencies, darker shades of gray depict shorter latencies. Observed data points are represented by asterisks.

When the head is not included in the strongly connected component, the main trends are inhibition from the secondary family size for compounds with log head family size less than 2, and facilitation from head family size for compounds with (residualized) log secondary family size roughly in the interval (-1, 1). Note that there are no data points in the lower right quadrant of the plot: large head constituent families invariably give rise to large secondary families.

The clear inhibition from secondary family size for small head family sizes follows from the (simplifying) assumption that when the family is large, a unit of activation is divided equally across all family members. Since the amount of activation spreading to secondary family members is smaller when the family size is large, the amount of noise contributed by secondary family members is greater for smaller families.

When the head is part of the strongly connected component, we find general facilitation for head family size. Interestingly, compounds with residualized log secondary family sizes around zero elicited the shortest latencies, now across the full range of head family sizes. For larger secondary families, we observe inhibition just as for compounds that are not part of the strongly connected component, giving rise to the longest (fitted) latencies. For smaller secondary family sizes, there is a hint of inhibition. Note that compounds in the strongly connected component with small head families and small secondary families are rare.

In this analysis, the factor specifying whether the head is part of the strongly connected component can be replaced by a factor specifying whether the head constitutes an exception to acyclicity. The resulting model is very similar to the one obtained on the basis of membership in the strongly connected component, but the fit is slightly less tight ($F(5.8699, 1227.8588) = 3.5056, p = 0.0021$, however, as the models compared are not nested, this likelihood ratio test is informal). In the following analyses, we therefore proceed with scrutinizing the tightly connected component. Irrespective of which factor is used, a simple main effect does not reach significance, indicating that there is no overall processing advantage to being exceptional or being part of the strongly connected component.

Figure 4 presents the surfaces fitted to the naming latencies, for which the same three-way interaction reached significance. For compounds with heads outside the strongly connected component, we have the same general pattern as for lexical decision. For compounds with heads in the strongly connected component, we observe mainly an effect of head family size. Effect sizes are smaller compared to lexical decision (contour lines are 0.02 log units apart in the contour plots for both tasks, but in the plots for lexical decision there are more contour lines that are closer to each other).

Both naming and lexical decision suggest that large secondary family sizes slow lexical processing. It follows that lexical decisions are not based just on the aggregated amount of activation in the mental lexicon, with more activation allowing shorter response latencies. If *tea-trolley* is presented, *trolley* may activate *bus stop*, but this compound is more likely to slow the lexical decision instead of speeding it. Apparently, lexical decisions involve discrimination between semantically relevant and semantically irrelevant lexical activation.

The presence of inhibition from the secondary family size in the naming task shows that this task is also sensitive to irrelevant lexical activation. Since *tea-trolley* is named

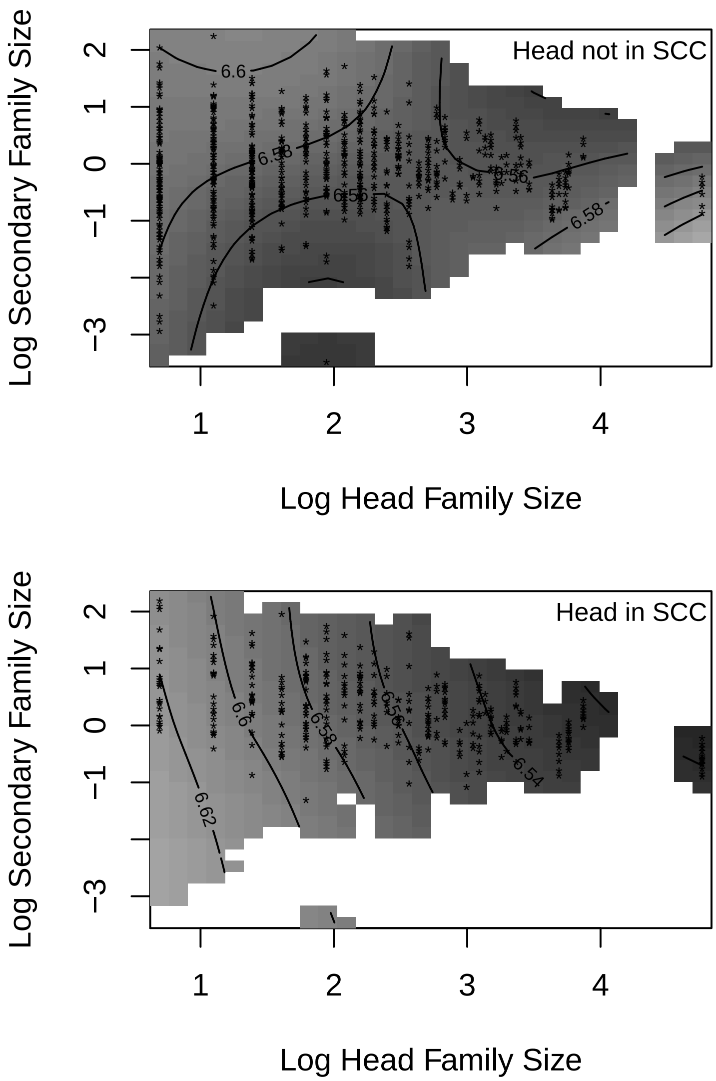

Figure 4: Fitted naming latencies in the English Lexicon Project as a function of log head family size and Residualized log secondary family size, for compounds not in (upper panel) and in (lower panel) the strongly connected component, fitted by a generalized additive model. Contour lines connect identical predicted log latencies. Lighter shades of gray represent longer latencies, darker shades of gray depict shorter latencies. Observed data points are represented by asterisks.

slower due to the co-activation of *bus stop*, the mediated priming effect observed by Balota and Lorch (1986) may be specific to the priming task. In unprimed contexts, activation of semantically too distant neighbors is probably detrimental. If this line of reasoning is correct, generalization from priming to normal (unprimed) processing is hazardous.

The absence of inhibition in the naming task for larger secondary family sizes for compounds in the strongly connected component is surprising. In fact, if anything, there is facilitation from larger secondary families, instead of inhibition. As compounds in the strongly connected component generate more co-activation, due to greater connectivity, mechanisms for ignoring irrelevant co-activation must be in place anyway for selecting the correct target for articulation. Possibly, the co-activated secondary family members are rendered harmless by the same deactivation mechanism.

This line of reasoning would be strengthened if it could be shown that connectivity in the strongly connected component can indeed be detrimental to lexical processing. We explore this possibility by means of the shortest paths from head to modifier in the strongly connected component.

4 Shortest Paths in the Compound Graph

In a study of mediated priming, Balota and Lorch (1986) observed that in word naming a word such as *cat* can prime *taxi* thanks to the mediating word *cab*, which is a form neighbor of the prime and a semantic neighbor of the target. Given that activation spreads beyond immediately related words, the question arises of whether activation likewise spreads within the directed compound graph from the head to the modifier. If so, there are two possible consequences for lexical processing. Activation spreading back from the head to the modifier might strengthen the activation of the modifier, facilitating lexical processing. Alternatively, given the inhibition observed from the secondary family size, activating the modifier from a chain initiated by the head might create uncertainty about which constituent is head and which constituent is modifier, leading to longer processing times: when *worm* in *wormwood* receives activation from the chain *woodcock* → *cockhorse* →*horsehair* →*hairoil* →*oilsilk* →*silkworm*, it is (re)activated as a head, while it functions as a modifier in *wormwood*.

To explore these possibilities, we calculated for each head in the strongly connected component the shortest path from the head to the modifier, using the `johnson.all.pairs.sp` function in the `RBGL` package (Carey et al., 2009). Examples of cycles illustrating shortest paths are shown in Figure 5.

The shortest paths show a skewed distribution with a long right tail. Most of this skew is removed by a logarithmic transformation. A generalized additive model fitted to the naming latencies revealed an effect of shortest path length, in interaction with log modifier family size, as illustrated in the bottom left panel of Figure 6. This interaction was modeled with a tensor product ($F(8.048, 424.453) = 6.224, p < 0.0001$) that outperformed simpler models with separate splines for modifier family size and shortest path length ($F(2.5251, 427.9786) = 3.2529, p = 0.02883$). The upper right panel shows the tensor product smooth for the interaction of head family size and secondary family size

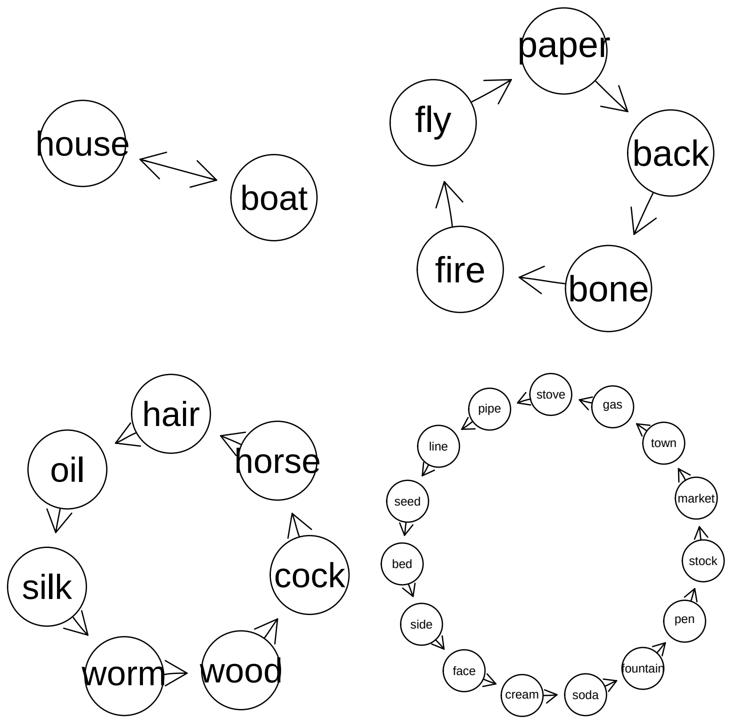

Figure 5: Examples of cycles in the compound directed graph: shortest head-to-modifier paths for *boat*→*house, back*→*paper, worm*→*silk*, and *stove*→*gas*.

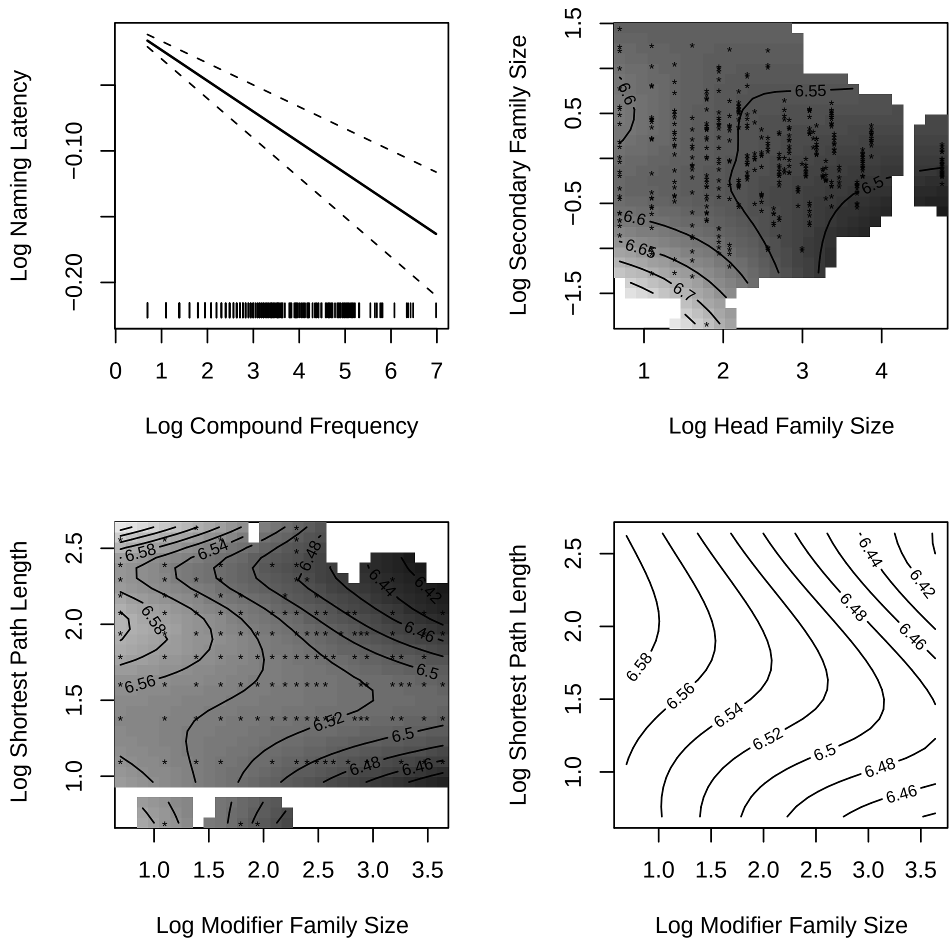

Figure 6: Partial effects in a generalized additive model fitted to the log naming latencies for the words in the strongly connected subgraph. Effects are shown relative to the intercept. Darker shades of gray indicate shorter naming latencies. The lower right panel shows the functional approximation of the panel to its left by equation (1).

($F(8.499, 424.453) = 6.224, p < 0.0001$), and the upper left panel the facilitatory effect of compound frequency ($t(424.453) = -6.934, p < 0.0001$), completing the full specification of the generalized additive model fitted to the data.

The contour plot in the lower left of Figure 6 shows, first of all, a facilitatory effect of modifier family size: we find darker shades of gray to the right. Furthermore, there is a ridge of higher naming latencies for intermediate shortest path lengths. For the smallest modifier families, the crest of this ridge is around log shortest path lengths of 2. As the modifier family size increases, this crest moves to slightly smaller shortest path lengths.

To understand the latency surface for modifier family size and shortest path length, we proceed from the assumption that activation spreading from the head to the modifier is disruptive, creating ambiguity about the functional status of the first constituent. The first constituent is a modifier, but at the end of the cycle receives activation from another modifier, suggesting it is a head. This disrupts the process of compound interpretation which, as shown by Gagné and Shoben (1997); Gagné (2001); Gagné et al. (2005), is driven by the distribution of conceptual relations instantiated in the compounds in the modifier family.

Across the range of values for modifier family size, we observe an initial increase followed by a decrease in latency as the shortest path length is increased. This suggests that there is a critical time window during which the incoming activation is especially disruptive. As the number of links in a cycle increases, the amount of time required for activation from the modifier to spread to the head, and from the head back to the modifier, increases as well. For very short shortest path lengths, the incoming spreading activation arrives to early to affect the process of compound interpretation. For very long shortest path lengths, it arrives too late.

The fitted surface is characterized by a ridge, extending from approximately (0.69, 2) to (3.5, 1.5) roughly along a straight line. This linear relation follows from the assumptions (i) that the amount of time required for activation to reach the modifier is proportional to the (log) shortest path length (L), and (ii) that the amount of time for sufficient activation to accumulate to be disruptive is proportional to (log) modifier family size (F). For larger families, activation spreads out more thinly. More time is required for sufficient activation to build up. To obtain an equivalent amount of disruption, a shorter shortest path length is required.

Formally, we can approximate the fitted surface obtained with the tensor product (a mathematical black box) with the following explicit parametric function for naming latency T (time), with as arguments modifier family size F and shortest path length L,

$$T(F, L) = a - b(F - A \sin[\omega(L - \phi)]), \tag{1}$$

and with parameters

$$a = 6.6065 \quad \text{(intercept)} \tag{2}$$
$$b = 0.0457 \quad \text{(slope for modifier family size)} \tag{3}$$
$$A = 0.2407 + 0.1847F \quad \text{(amplitude of the "ridge" sine)} \tag{4}$$
$$\omega = 2.4867 \quad \text{(angular frequency of "ridge" sine)} \tag{5}$$
$$\phi = 1.6664 - 0.2610F \quad \text{(phase shift of "ridge" sine).} \tag{6}$$

Parameter values are estimated from the fitted partial effect of modifier family size and shortest path length shown in the lower left panel of Figure 6, using mean squared error minimization by means of a nonlinear conjugate gradient method (Fletcher, 1987) on a 30 by 30 grid (MSE= 0.0003 at convergence).[1] The resulting approximated surface is shown in the lower right panel of Figure 6.

The general negative slope for the modifier family size is represented by b. On a given contour line, T is constant, in which case $F = A\sin[\omega(L-\phi])$, yielding a "ridge" sine for F as a function of L. The amplitude of this ridge sine is modeled as linear in F: as modifier family size increases, the amplitude increases. Finally, the phase shift ϕ of the ridge sine decreases linearly with F. The crest of the ridge is reached when $\omega(L - \phi) = \pi/2$. By (6), this is equivalent to $L + 0.2610F = \pi/(2\omega) + 1.6664$, which is a constant independent of L and F. Thus, increasing L implies decreasing F and vice versa, consistent with the interpretation of the location of the crest as the statistical fingerprint of a critical window in time in which activation reaching the modifier interferes with the interpretation of the compound, slowing the naming latencies.

The lexical decision latencies did not reveal any effect for log shortest path length. The presence of an effect of shortest path length in word naming and its absence in lexical decision fits well with the results of Balota and Lorch (1986), who observed mediated priming in word naming but not in lexical decision. Balota and Lorch attribute the absence of an effect in the lexical decision task to a post-access verification stage specific to lexical decision, during which subjects would execute a lexical decision conditional on having checked whether there is a semantic relation between primes such as *cat* and targets such as *taxi*. As the data from the English Lexicon Project are from unprimed lexical decision, a more general explanation is called for. The crucial difference between naming and lexical decision is that in naming a specific compound has to be selected for articulation, while a lexical decision can be based on the amount of lexical activation triggered by the visual stimulus. If a decision is based, at least in part, on this general lexical activation (Grainger and Jacobs, 1996), then the small effect of mediated activation due to head-to-modifier cycles may be washed out by the much larger activation contributed by a word's morphological family members.

[1] I am indebted to Jorn Baayen for his help with formulating and fitting this model.

5 General Discussion

This study explored the connectivity in the lexical network of English compounds with conceptual tools from graph theory, and studied the consequences of this connectivity using the naming and decision latencies available in the English Lexicon Project. The compound directed graph, although not acyclic, revealed the same surprising tendency towards acyclicity that characterizes suffixation (Hay and Plag, 2004; Plag and Baayen, 2009) and prefixation (Zirkel, 2010) in English. As for suffix sequences, we observed a positive correlation between productivity and rank for heads. For modifiers, we observed a negative correlation between productivity and rank, reflecting the results for prefixation. Rank (or Y coordinate) in the DOT representation of the directed graph did not enter into any further correlations with distributional measures of processing complexity (relative frequency, junctural biphone frequency), nor with naming or lexical decision latencies. For compounds, the conclusion is that the significant tendency to acyclicity cannot be derived from principles of processing complexity.

This conclusion raises the question of why the tendency towards acyclicity exists. Plag and Baayen (2009) speculated that acyclicity affords enhanced prediction of upcoming constituents. In this study, we explored the complementary possibility that extensive connectivity might have adverse effects on lexical processing due to activation spreading to irrelevant words. It is well known that greater primary connectivity, as measured by the positional family size of head and modifier, goes hand in hand with shorter processing latencies (De Jong et al., 2002; Kuperman et al., 2008, 2009). It turns out, however, that the count of all family members of the compounds' constituents, the secondary family size, has an inhibitory effect. Especially for heads with a small primary family, larger secondary families give rise to elongated latencies in lexical decision and word naming. In the lexical decision task, the inhibitory effect of the secondary family size is especially prominent in the strongly connected component of the compound graph, the part of the graph where connectivity is most dense, and where activation of irrelevant words resonates most strongly.

The adverse effects of the co-activation of more distant morphological relatives was explored further by investigating the shortest path lengths from head to modifier for compounds in the strongly connected component. Delayed naming latencies emerged for intermediate shortest path lengths. The path length with maximum inhibition decreased for increasing modifier family size. This pattern was modeled as reflecting a critical time window for interference from activation arriving back at the modifier to interfere with the interpretation of the semantic relation between modifier and head.

One of the issues raised by Balota and Lorch (1986) is how, given the massive spreading of activation into the lexical network, individual words can still be identified. The present study provides a partial solution by demonstrating that not all connectivity is helpful. Apparently, the algorithms used to meet the requirements of the naming and lexical decision tasks are sensitive to the semantic relevance of co-activated words. In the naming task, which requires a unique response, the de-activating of irrelevant more distant morphological relatives is seen most prominently: the effect sizes of the secondary family size are much reduced in this task compared to lexical decision.

The present results challenge the usefulness of the priming paradigm as a tool for understanding normal lexical processing. The primed naming task used by Balota and Lorch (1986) revealed an effect of mediated priming (from *cat* through *cab* to *taxi*). In our naming data, the effect of the secondary family size is inhibitory, instead of the facilitatory effect one would expect given the immediate priming results. The present data lead to the prediction that in unprimed naming, the processing of *taxi* is delayed by the mediated co-activation of *cat*.

We end this study with a cautionary note. The analyses presented here are all exploratory in nature. A wide variety of measures were explored, and only the predictors that turned out to have robust explanatory value are reported. Therefore, replication studies for larger data sets, and different languages, will be required before the present results can be established as more than a promising window on the pros and cons of morphological connectivity in the mental lexicon.

References

Arndt-Lappe, S. (2010). Towards an exemplar-based model of English compound stress. *submitted*.

Baayen, R. H., Kuperman, V., and Bertram, R. (2010). Frequency effects in compound processing. In Scalise, S. and Vogel, I., editors, *Compounding*. Benjamins, Amsterdam/Philadelphia.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Balota, D. and Lorch, R. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(3):336–345.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3):445–459.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics. Identifying Influential Data and sources of Collinearity*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Carey, V., Long, L., and Gentleman, R. (2009). *RBGL: An interface to the BOOST graph library*. R package version 1.18.0.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects. *Brain and Language*, 81:555–567.

Fletcher, R. (1987). *Practical Methods of Optimization. 1987.* John and Sons, Chichester.

Gagné, C. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27:236–254.

Gagné, C. and Shoben, E. J. (1997). The influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23:71–87.

Gagné, C., Spalding, T., and Ji, H. (2005). Re-examining evidence for the use of independent relational representations during conceptual combination. *Journal of Memory and Language*, 53(3):445–455.

Gansner, E. R., Koutsofios, E., North, S. C., and Vo, K.-P. (1993). A technique for drawing directed graphs. *IEEE Transactions on Software Engineering*, 19(3):214–230.

Gentry, J., Long, L., Gentleman, R., Falcon, S., Hahne, F., and Sarkar, D. (2009). *Rgraphviz: Provides plotting capabilities for R graph objects*. R package version 1.20.3.

Grainger, J. and Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103:518–565.

Harrell, F. (2001). *Regression modeling strategies*. Springer, Berlin.

Hay, J. B. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, 39:1041–1070.

Hay, J. B. (2002). From speech perception to morphology: Affix-ordering revisited. *Language*, 78:527–555.

Hay, J. B. and Plag, I. (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.

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics*, 39(1):51–93.

Krott, A., Hagoort, P., and Baayen, R. H. (2004). Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds. *Language and Cognitive Processes*, 19:453–471.

Krott, A., Schreuder, R., Baayen, R., and Dressler, W. (2007). Analogical effects on linking elements in german compounds. *Language and Cognitive Processes*, 22:25–57.

Kuperman, V., Bertram, R., and Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23:1089–1132.

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2009). Reading of multimorphemic Dutch compounds: towards a multiple route model of lexical processing. *Journal of Experimental Psychology: HPP*, 35:876–895.

Plag, I. (2010). Compound stress assignment by analogy: the consituent family bias. *Zeitschrift für Sprachwissenschaft*, 29.2.

Plag, I. and Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language*, 85:106–149.

Plag, I. and Kunter, G. (2010). Constituent family size and compound stress assignment in English. *Linguistische Berichte*, Sonderheft 17.

Plag, I., Kunter, G., and Lappe, S. (2007). Testing hypotheses about compound stress assignment in english: a corpus-based investigation. *Corpus Linguistics and Linguistic Theory*, 3:199–232.

Wood, S. N. (2006). *Generalized Additive Models*. Chapman & Hall/CRC, New York.

Zirkel, L. (2010). Prefix combinations in English: Structural and processing factors. *Morphology*, in press.